THE SUPPLY AND DEMAND FOR ELECTRICITY IN AUSTRALIA

(A theoretical and empirical study for the years 1953 to 1971)

by

T. WEYMAN-JONES

(A thesis submitted for the degree of Ph.D. in the University of Melbourne)

THE SUPPLY AND DEMAND FOR ELECTRICITY
IN AUSTRALIA

CONTENTS

PART III

ESTIMATING THE DEMAND FOR ELECTRICITY

PART IV

CONCLUSION

FOREWORD

This thesis is submitted for the Ph.D. degree in the Faculty of Economics and Commerce, at the University of Melbourne. The work has been supervised in turn by Professors W. Prest, C.S. Soper and L.R. Webb. I would very much like to thank each for the help and encouragement I received. Roy Webb took a particular interest in the work at an important stage and the whole of part II follows a direction pointed out by him.

In addition, various drafts of different parts of the thesis have been read in detail by Brian Parmenter and Peter Jonson and both have been very helpful in providing critical stimulus. Brian Parmenter proved a particularly demanding and constructive critic of part II of the thesis.

The Electricity Council of the U.K. and the Electricity Supply Association of Australia both provided library facilities and I would especially like to thank Mr. Kay of the Melbourne office of the E.S.A.A. for his help in locating data.

Carol Dawson has done an excellent job of preparing the typescript. The bulk of the work was carried out with the help of a Commonwealth Scholarship, for which I am grateful.

My wife, Carol Weyman-Jones, has been a constant source of encouragement to me and deserves the largest acknowledgement I can give.

This study is my own original work and none of the above is responsible for any errors remaining in it.

LOUGHBOROUGH UNIVERSITY
OF TECHNOLOGY.
MAY, 1975.

T. WEYMAN-JONES

SUMMARY (IN FULFILMENT OF REGULATION 3.60)

This thesis considers some theoretical aspects of supply planning and undertakes empirical analyses of both demand and supply behaviour in the Australian electricity industry over the period 1953 to 1971.

The research strategy reflects the way in which an economic adviser to an electricity authority might organize his work. The initial step is to spotlight three important areas for analysis: the link between demand and supply provided by cost and price signals; the planning of supply capacity to meet demand targets; and, finally, the forecasting and empirical analysis of demand.

Part I of the thesis considers a traditionally important field of electricity economics - optimal pricing. Several tariff schemes suggested by economists and engineers are surveyed in conjunction with the reasoning behind them. Extensive tariff data, especially collected for the thesis, is then compared with reported costs (roughly in running and capital cost categories) to discover the extent to which theoretical tariff principles have been observable in practice. The empirical results suggest an imperfect use of theoretical tariff principles. This finding clearly prompts several questions: e.g. are there significant costs in non-optimal tariff formulation; is the peak loading problem sensitive <u>in practice</u> to tariff manipulation?

The empirical studies carried out in Chapters IV and V suggest there are significant measurable costs of resource misallocation and that peaks in the load could well be significantly varied by tariff manipulation.

Part II of the thesis is a detailed theoretical analysis of investment planning in electricity supply to meet specified demand targets at minimum total system cost in present worth terms. This analysis is set in the context of the celebrated model of Ralph Turvey (see bibliography); his principal models are examined and criticised; a survey of the background literature is given beginning with the classic article of Hotelling on depreciation. Subsequent chapters present a practical formulation of an electricity capacity planning problem for the Australian situation and a detailed survey and critique of the idea of "dynamic marginal cost".

Part III of the thesis turns to the demand side. A thorough empirical study (using a variety of models of demand) is carried out for the residential sector and it suggests the usefulness of a model that simultaneously models the demand for the services of a stock of consumer durables and the demand for electricity as an input to the process of generating those services. Attention is also paid to traditional methods of measuring price elasticity of electricity demand. Finally, the industrial and commercial sectors come under scrutiny and the tariff data especially collected for the experiments reported in part I are used to test and compare conventional "average revenue" and the more theoretically correct "marginal price" methods for calculating price elasticity of demand.

This concludes the study. A more detailed summary appears as Chapter XIV of the thesis.

CHAPTER I

ELECTRICITY IN AUSTRALIA

This is a study of the supply and demand for electricity in Australia by an outside observer. The motive for the study is academic and the constant if not the only emphasis is on the application of theoretical and applied economics. Several other approaches might have been adopted. For example, detailed study of energy policy options and the cost of alternative primary energy inputs is only in its infancy in Australia; however to have followed this path, in the present state of published statistics, would have put some strain on the usual economic calculations. Where such work is being done for other economies (for example the U.S. and the U.K.), it has been preceded by more analytical studies well within the competence of the outside observer and by empirical work that can usefully adopt published data.

The theme of the study is to apply to the Australian electricity industry some of the advances in theoretical and empirical electricity economics that have appeared in recent years. Nevertheless, in many places, particularly in part I of the thesis, completely new approaches are adopted to understand the economics of Australian electricity. Like other developing countries, data for the Australian electricity industry is relatively rich on the sales and tariff aspects, but relatively rare on the details of investment choice. Consequently the bulk of the empirical work in this study is related to demand and price movements, while detailed investment programme analysis receives a much more theoretical treatment. This pattern of development of critical work mirrors that in other countries.

STRUCTURE OF THE STUDY

The remainder of this first chapter will present the main facts about the organization of the Australian electricity industry and the reasons for taking up the different methods of analysis that will be used. After this, the thesis is divided into three parts, each of which reflects an aspect of electricity economics which can satisfactorily be treated by an academic economist. To a certain extent the dividing lines used evolved from the way in which an electricity authority's economic adviser might begin his job. The first statistics that an authority publishes are about its financial operations: revenues, accounting costs, investment undertaken, tariffs charged. (The historical development of electricity statistics is readily seen from the admirable "Statistics of the Electric Supply Industry" published by the Electricity Supply Association of Australia). An imaginary academic economist entering the industry has first to work with these statistics. He will be at the same time concerned with the interdependence between different areas of the authority's operations, but will quickly recognize Ralph Turvey's point about the necessity of adopting a piecemeal approach to the analysis and later iterating between one section and another[1].

On this basis, part I of the thesis looks at Tariff policy in detail. Electricity economics began as an attack on the peak load pricing problem and the design of tariffs has been its longest pre-occupation. The strategy of part I of the thesis is to look at the tariff setting problem with as broad a view as possible - to include economists' and non-economists' suggestions and to examine the implications for resource allocation. Here the Australian data is richest and, while no more than a beginning is attempted, this part of the thesis does try to go as far and as widely as possible in examining tariff and cost relationships in Australian electricity. (This part of the thesis would not have been possible without the co-operation of many of the Australian electricity authorities in providing tariff schedules in some cases going back to 1942.)

Having established some tariff information, the imaginary economic adviser - again following Turvey - faces up to two fundamental problems: the existing tariffs (and other economic factors) will produce a set of demand targets over the coming years; how are these targets to be achieved and how is demand to be forecast? Once more the inter-dependences have to be temporarily set aside and the problems tackled iteratively. Part II of the thesis (which begins with a form of flow chart for this sectioning of the economists' problems), tackles the problem of the investment programmes to meet specified output targets. This part of the thesis remains theoretical for two reasons; first of all it is written by an outsider rather than a particular adviser and so detailed and confidential data on plant running is not available; in any case each state would have to be treated separately. Secondly, and more importantly, the analytical problems of choice of technique and investment programming analysis must first be solved and this is an area of theoretical ferment at present. Part II of the thesis therefore concentrates on analytical issues in planning investment policy, though chapter VIII does present a model which might immediately be used by an Australian generating authority.

Returning to the imaginary economic adviser who has now set up his investment programming model, the next hurdle he meets arises from the fact that his investment programme yields measures of marginal cost to which tariffs might be related in some way. (Exactly what marginal costs and how they are calculated is another theme of part II.) If tariffs are then based on these up-to-date costs (marginal or average or whatever) there may very well be a change in demand and the emergence of a new set of demand targets. The imaginary economic adviser therefore must see as his third task, the forecasting and analysis of demand. This is the concern of part III of the thesis, which is a study, in the tradition of the econometrics of demand, of the various sales sectors for electricity. The aim is to fit standard demand equations of the sort that may yield the elasticity estimates essential for forecasting demand response to different tariff and cost calculations.

Thus, while the thesis must remain purely academic, I have tried to follow the path which a newly appointed economic adviser might take in an electricity authority. The object is to produce one of the many initial studies necessary for detailed evaluation of Australian energy policy. It begins now with some of the factual and organizational background to the Australian electricity industry.

ELECTRICITY SUPPLY INDUSTRY

The aim of this section is to provide a general background to some of the issues that will be discussed in this study. The electricity industry in all developed industrial nations tends to evolve with the economy itself: there is a large expansion in the demand for electricity as an industrial input as industrial production expands and there is a further large expansion in residential demand for electricity as the demand for appliances rises with rising living standards. This sort of expansion has characterised the post-war growth of the Australian electricity industry though, as Prest points out, there was an earlier phase of growth when electricity replaced other fuels in transport and public lighting needs (Prest (1963)).

Table I, 1 presents some figures on the growth of demand over the post-war period. In the decade of the fifties overall consumption increased more than two and a half times and in the sixties it was almost exactly doubled (equivalent to an annual compound growth rate of about 7.2%). Over the decade of the sixties residential consumption accounted for a stable 38-40% of this total consumption. To meet this rising demand the electricity authorities must undertake massive capital investment programmes. Table I, 2 indicates, for approximately the same decades, the growth in installed generating capacity. It is interesting that the growth rates here are even greater. Over the decade of the sixties generating capacity rose by a factor of two and a half. These are the sort of trends that highlight one of the perennial problems of electricity generating authorities: the problem of peak loads.

Table I, 1 Electricity Consumption in Australia[2]

(billion Kwh)

1951-52	1955-56	1961-62	1965-66	1971-72
10.0	15.3	27.0	38.0	55.9

Table I, 2 Generating Plant Installed[3] (GW)

1949-50	1959-60	1969-70
2.4	5.6	14.0

Electricity cannot be easily stored and consequently has to be produced as it is demanded. Enough generating capacity to meet maximum demand must be installed if supply restrictions are to be avoided. However when demand becomes more peaked the result is that more and more capacity has to be installed in order to meet a fraction of annual consumption and to stand unused for the remainder of the year. In fact the situation is slightly more complex since newly installed capacity usually has running cost advantages over older capacity which ensure that new capacity always meets base load; increased concentration of consumption at peaks then means that older capacity is moved more rapidly into the stand-by sections of the "merit-order" running system. The basic consequence remains that the more peaked is demand the lower the load factor of the authority's operations.

Peaks in demand impose additional costs in terms of using up scarce resources. This is not an insoluble problem and building more and more capacity is not the only answer. It is not often realised that an alternative to capacity expansion to meet maximum demand growth is a restructuring of tariffs to bring home marginal resource costs more clearly to consumers.

This growth of the peak load problem has had an impact on the electricity authorities in recent years. At the national level, "Energy in Australia" the Department of National Development's Survey published in 1967 commented (page 24)

"... Short duration peak loads, particularly winter evening peaks have grown at a very rapid rate indeed over the past few years ... The extraordinary growth of peak demands means that the utilisation time or load factor of the plant is falling ... "

The use of off-peak tariffs to spread this load has had only qualified support though the State Electricity Commission of Victoria reported in 1971 that

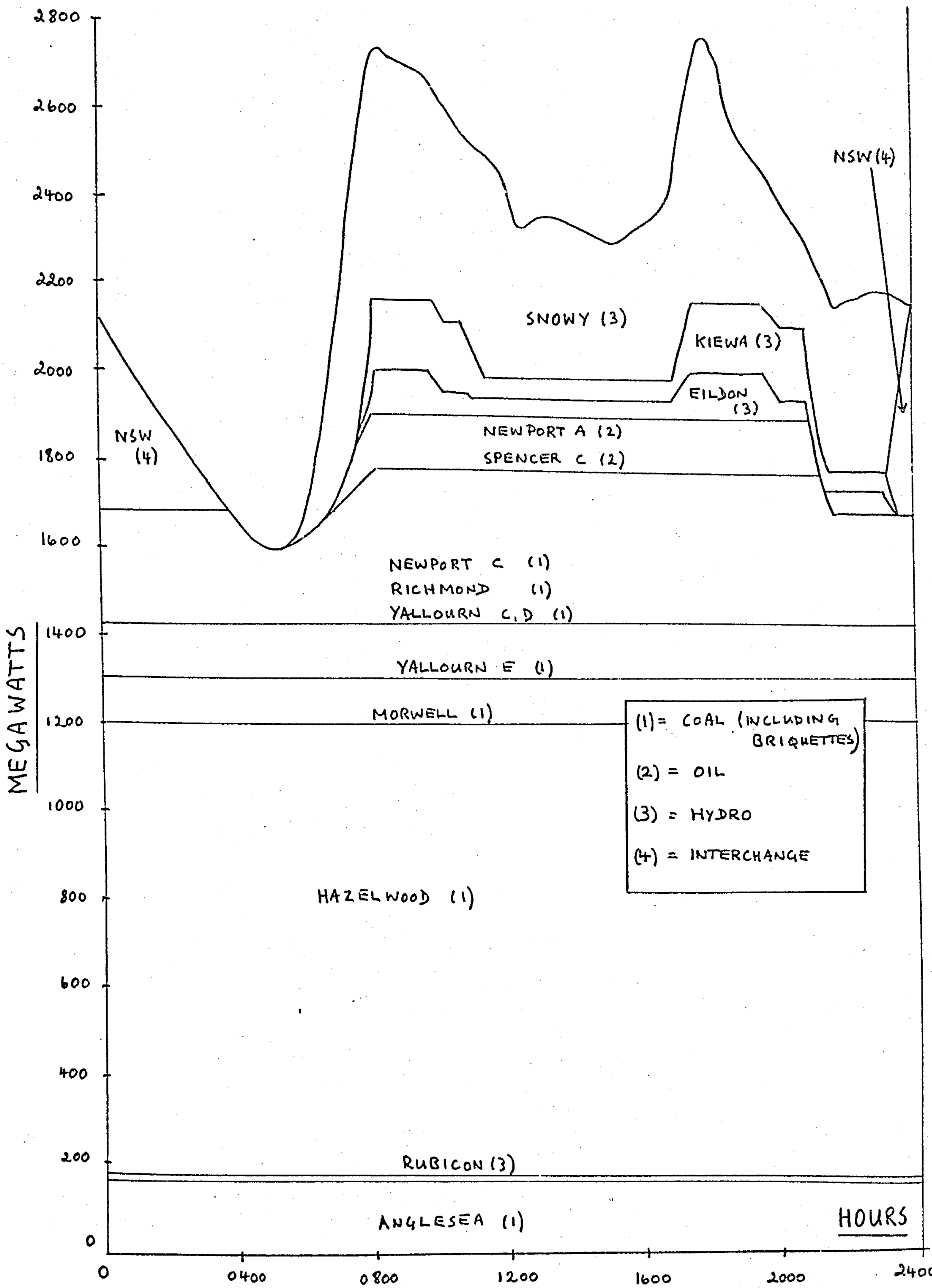
"... marketing activity was intensified with emphasis on the promotion of off-peak electricity sales ... "

(Annual Report (1970-71) page 14, State Electricity Commission of Victoria)

Figure I, 1 shows the extent of the SECV's problem: it is a reproduction of the Load Curve for the 1971 day of maximum demand. The other authorities have very similar patterns of daily fluctuations.

This whole question of the costs of peak demands and the use of tariffs is tackled in detail in a series of empirical studies in part I of the thesis.

FIGURE I, 1


2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
MEGAWATTS
NSW (4)
NSW (4)
SNOWY (3)
KIEWA (3)
EILDON (3)
NEWPORT A (2)
SPENCER C (2)
NEWPORT C (1)
RICHMOND (1)
YALLOURN C, D (1)
YALLOURN E (1)
MORWELL (1)
(1) = COAL (INCLUDING BRIQUETTES)
(2) = OIL
(3) = HYDRO
(4) = INTERCHANGE
HAZELWOOD (1)
RUBICON (3)
ANGLESEA (1)
HOURS
0
0400
0800
1200
1600
2000
2400

Aside from the problem of peak loads, there are other characteristic aspects of the electricity industry. One of these is the choice of fuel inputs to use in the generating system. There are five candidates for fuel choice in most advanced systems: coal, oil, natural gas, hydro and nuclear. Only the last is not currently used in Australia though recent years have seen some discussion of the possible date for introducing nuclear power.

The extent of fuel mix is apparent when Figure I, 1 is examined again. For the case of Victoria base load is met almost entirely with the brown coal steam raising plants of the Latrobe Valley (Hazlewood, Morwell and Yallourn). As the peak builds up oil fired plants are called on until finally peak demand is met by the Eildon, Kiewa and Snowy hydro schemes. The Snowy Mountains hydro electricity partly comes from an interchange with New South Wales where it also meets the peak which does not coincide with the Victorian peak.

In New South Wales, coal - black coal this time - again meets base load and a proportion of peak load; oil and hydro supplement this. The hydro entitlement from Snowy of the NSW Electricity Commission is double that of Victoria and in 1971-72 this amounted to nearly fourteen percent of total NSW consumption.

Queensland uses black coal almost exclusively, though a very small amount of power is generated from steam raised by natural gas. Most of the planned capacity expansion is for coal using thermal equipment.

Tasmania of course is almost exclusively a user of hydro electricity - it has about half of total Australian hydro potential and capacity expansion is still envisaged in hydro schemes. These have incurred the wrath of conservationists because many of the proposed works may destroy areas of nature reserve. This is obviously a question susceptible of economic analysis since it could be argued that load growth can be diverted by corrected shadow pricing. What is needed is a measure of the extent to which this can be done and the cost savings involved. Chapter IV of this thesis will attempt this.

South Australia has concentrated in the past on the use of sub bituminous black coal chiefly from the Leigh Creek coalfield, but discoveries of natural gas raised the possibility of using this as a primary input. This is not usually an economic choice where (a) there are large coal reserves, (b) transmission losses from coalfield sites and water cooling problems on coalfield sites do not make coal an uneconomic choice of input, because natural gas has a large opportunity cost when used as a bulk industrial fuel. South Australia's coal position is such as to make natural gas an economic choice of input and by 1971-72 natural gas burn in power stations had risen from zero three years previously to 24 billion cubic feet.

West Australia uses black coal as its chief fuel, partly imported and partly indigenous.

The plant mix problems presented by these fuel choices are of three kinds:

(i) What is the optimal combination of plant type and fuel use bearing in mind that plants differ in capacity costs, running costs and maintenance needs?

(ii) Are there budgetary limits on an authority's operations which mean that the monetary cost of a plant with high capacity installation charges is below its scarce resource shadow cost? Discount rates have a role to play here.

(iii) What is the implication for plant choice of the fact that some of these inputs are exhaustible resources (particularly hydro sites and gas reserves)? How does this have a bearing on the decision date for the introduction of nuclear power?

Such questions can only be answered by detailed investment analysis of the sort carried out in part II of this thesis. The choices taken in practice will have an effect on costs incurred by the authorities and hence on the tariffs charged.

A final characteristic of the electricity supply industry which is important from an economic point of view is the nature of its organization. Electricity supply organizations today are usually State owned or run as State regulated public utilities. The Australian pattern is for ownership to be in the hands of a State-wide authority; these differ in their organizational details (for a good guide see Prest (1963)) but tend on the whole to be based on the British pattern whereby the generating authority is a public commission, with tariff making, distribution and marketing operations carried out either by the same or similar state bodies or by local councils using the generating authority's bulk supplies.

There are strategic, national interest and economies of scale arguments for state ownership of the generation side; the chief organizations are set out in Table I, 3 with the other regulatory bodies involved.

The authorities have instructions not to interrupt supplies, and "Energy in Australia" (page 22) mentions the avoidance of supply restrictions as one of the chief headaches of the early post-war era. Beyond this there is no committment to an economically determined pricing or investment policy; there is even no explicit suggestion from State or Commonwealth government about the appropriate use and choice of a discount rate. In developing Australian energy policy this is the largest gap to be filled. Nevertheless each authority has borrowing committments which compel it to work closely with the appropriate State Treasury. State Treasury loans, semi governmental bond issues and internal financing have all provided capital funds for expansion (Prest (1963), page 34) and it is probable that some form of resource allocation analysis takes place in the States' budgetary negotiations with their authorities. In this context, whether or not used as a price basis, marginal cost calculations are explicitly or implicitly made. The detailed calculation of marginal resource costs cannot be avoided by any authority and the long life of its

Table I, 3 Organisation of the Electricity Industry

STATE	ORGANISATION	FUNCTIONS
NSW	Electricity Commission	Generation and bulk supply to distributing authorities.
NSW	Electricity Authority	Regulatory body.
VICTORIA	State Electricity Commission	Generation and regulation, tariff setting etc.
QUEENSLAND	State Electricity Commission	Generation and bulk supply.
	Southern Electricity Authority, Northern Electricity Authority, Brisbane City Council, and others	Operation of distribution systems.
S.A.	Electricity Trust	Generation and operation of system.
W.A.	State Electricity Commission	Generation and regulation of supply.
TASMANIA	Hydro Electric Commission	Sole electricity authority.
ACT	ACT Electricity Authority	Operation and regulation using power from Snowy Mountains Hydro-electric Authority (in conjunction with Snowy Mountains Council)
NT	Administered by Commonwealth Government Department	

capital equipment make it essential that some form of dynamic marginal cost model is used. Such models will be examined in part II of this thesis. More detail of the Australian electricity industry in its various State organizations will appear as the study develops.

NOTES TO CHAPTER I

1. Turvey (1968) and Turvey (1971a). The point is discussed several times in the chapters to come.

2. The source of the data is "Statistics of the Electric Supply Industry", (E.S.A.A.) various issues. The unit of measurement is the Kilowatt-hour (KWh) the standard measure of electric energy, measured as power over a period of time.

3. The source of the data is again "Statistics of the Electric Supply Industry" and the unit of measurement is the gigawatt (GW) equivalent to one million kilowatts, the standard unit of electric power provided at any one point of time.

PART I

A STUDY OF PRICES AND COSTS IN ELECTRICITY SUPPLY

CHAPTER II

PRICES AND COSTS IN ELECTRICITY SUPPLY

I would like to consider in detail an analysis of prices and costs in the electricity supply situation in Australia; the structure of the analysis will be

(A) - to consider what theoretical suggestions have been made for electricity pricing.

(B) - to set up an empirical measure by which the extent of the application of these suggestions can be gauged.

(C) - to consider what inferences may be made about costs of electricity supply.

(D) - finally to attempt some empirical analysis if possible of the crucial peak loading problem and the way it has been tackled in Australia.

(A) ELECTRICITY TARIFFS IN THEORY

Economic theory has suggested several ways of pricing electricity, all of which are practised in Australia at the present moment. Essentially, different tariff suggestions correspond to the different objectives being met in pricing policy.

A basic premise accepted by all of the schemes considered here is the absence of second best problems - Moreover electricity pricing has not been the province only of economists and the earliest attempts at it were developed by engineers. Very often the engineer's approach has been the subject of criticism but is is important to realise that the planner's objective will determine the eventual choice of a price setting mechanism. The economist's approach is traditionally a marginalist one based on some concept of social welfare. The engineer's objective is usually more precise and often is no more than recovering the full costs of the effort involved in satisfying the maximum demand on the system.

To begin with the economist's marginalist approach, consider a simple objective: minimizing costs of providing for a fixed set of demand requirements in the present and succeeding periods. An only slightly more complex objective is maximizing the different between willingness to pay and social cost; in this case if the planner feels that his problems are so complex that he can only tackle them piecemeal, then minimizing costs will still determine much of his operational effort. In this context the public enterprise economist is crucially interested in the price signals which he wants to send to consumers to advise them of the costs they impose on the electricity authority by any change in their plans. The price structure should signal these "marginal costs". Price signals are then acting as "the handmaiden of investment policy" (Posner 1973), because they tell consumers of the capacity and running costs associated with adjusting a given set of plant to cater for the caprices of their demands.

To develop this idea of price signals it is necessary to say something about costs and investment policy. At this stage I am only going to present some simple summary ideas of the way modern economics looks at these; these ideas are based on the more extended analysis of Part II of my thesis.

Modern theoretical approaches have stressed the dynamic nature of the planning problem with the result that the static traditional textbook ideas of short and long run marginal costs are not helpful; it is more useful to borrow some French concepts among which is the idea of the marginal costs associated with an "adjusted" system i.e. a system of plant and equipment which has evolved from following an optimal investment policy - (in this analysis optimal is synonomous with cost minimizing). All the structural changes to be considered are evaluated in terms of C* - the minimum value of the present worth of all system costs. The value of C* is the outcome of a policy of investment in plant in order to meet the forecast output requirements for each of the years of the planner's time horizon. The output requirements are represented by $(X_0, X_1, \ldots, X_t, \ldots)$.

Consider now the appearance of a new consumer who announces that he will require the following set of outputs beginning in year s of the plan: $(\Delta X_s, \Delta X_{s+1}, \ldots \Delta X_t, \ldots)$. The electricity authority can regard this as a <u>permanent output increment</u> and requires to find the cheapest way of meeting it. When this has been done and the plant system adjusted to meet these output increments in the least expensive way then the extra costs are signalled to the consumer in the form of prices. First, look at the characteristics of the existing system. There will exist at any time a set of plants of different ages and types for providing output and also a set of plans for plants for later years. The plants will generally differ in two ways:

(a) in the ratio of capacity to running costs; e.g. nuclear plant usually has a very high capacity cost and a very low running cost while at the other extreme small gas turbines have a relatively low capacity cost but a relatively high running cost. In between there is a spectrum ranging in turn through hydro facilities, oil and coal fired plants.

(b) The plants will also differ in the extent to which age has reduced their cost saving attributes. The most up-to-date plant of a given type will have lower running and capacity costs than older plant of the same type; the two forces of technological advance in design and manufacture and wear and tear due to use ensure this. The desirability of using any given plant (whatever its type and age) will depend on its "cost savings" - its ability to lower C* below what it otherwise would be. On the system at any time there are certain plants in operation and certain plants on standby. Those on the margin of operation could be called marginal plants and it is <u>their</u> running costs that are the <u>marginal</u> generating costs of the system. Any other plant <u>in</u> operation will show a running cost saving over the marginal plant and such plants will have been scheduled into the investment policy because the present worth of their lifetime cost savings was at least as great at the time of installation as their capacity costs.

In this context consider the permanent output increment the new customer brings along. He may have arrived at a planned or expected lull in demand, a period in which there is a large reserve of standby plant. For at least one period his demand is met by bringing in one of these standby plants at a cost that reflects only the marginal generating cost of the system. However eventually - perhaps in the next period - there will be no planned spare capacity and to meet his continuing extra demand the electricity authority has to build more capacity. It now incurs a stream of extra capacity and running costs. However the new capacity, if built, will not be kept in the reserve margin. Such will be its cost savings due to its "up-to-dateness" that it will push out some older plant into the reserve. The stream of running costs associated with the new demand will be reduced by having newer plant to deal with it, and the correct marginal cost term is the present worth of capacity costs and the running costs on the "adjusted" system of plant.

Of course if the consumer appears when there is no planned excess capacity at all then the new capacity must be built immediately and the marginal costs will be all the greater. These ideas are easily summarized:

(a) the consumer appears at a time period, θ, at which there is spare capacity. The total extra costs he imposes are

P.W. (θ, τ) of marginal generating costs

+

P.W. (τ, ∞) of capacity and marginal generating costs on the adjusted system.

(the expression P.W. (a, b) denotes the present worth now of an expenditure stream beginning at a and ending at b).

At the date τ there is no longer any spare capacity and new capacity is built.

(b) the consumer appears when there is no spare capacity at all and incurs costs of

P.W. (θ, ∞) of capacity and marginal generating costs on the adjusted system.

The present worth sum in case (a) must be lower than in case (b) and the economist will expect to see this signalled to the consumer. The electricity authority simply has to extract from the new customer a different lump sum of present worth incremental costs according to the starting date of the customer's demand. It is the overwhelming importance of the effect of different starting dates on costs which has to be signalled. Any way of extracting this sum will do. For instance the consumer may be offered one of several options; e.g.

(i) pay a lump sum for a contract of supply

(ii) pay an annuity over the years of supply

(iii) pay a varying annual sum declining from an initially high starting value

The customer will choose according to his discount rate and views of the future. Option (iii) is interesting because although it does not accord with the usual idea of discounting the future it does reflect what is happening to the electricity authority's costs. Remember that equipment is installed on the basis of projected lifetime cost savings. These cost savings can be expected to decline over the capacity's life due to the forces of technological progress and wear and tear. The choice of an "amortization" stream to recover capacity costs will then fall to a method that writes off more of the value of capacity in earlier years than in later years.

None of the above schemes makes any allowance for expected lulls in demand. In a period of large excess capacity the cost savings of some capacity may drop to zero only to rise again as demand picks up. To reflect this the authority may try a two part tariff:

(a) a capacity element that recovers installation costs according to the starting date of demand

(b) a running charge that varies with the level of excess capacity.

The chief point that emerges is that to the economist the starting date of the permanent output increment determines the cost of providing for it. This is the basis of what the French have called "dynamic marginal cost".

However this sort of cost recovery is not what electricity authorities have usually practised. Most of the tariff schemes in operation throughout the world do not recognize the permanent nature of a new item of demand. Instead new permanent customers are treated as a succession of instantaneous customers and some idea of "instantaneous dynamic marginal cost" needs to be developed by the economist. The most celebrated attempt to do this in the English literature is Ralph Turvey's 1969 paper in the Economic Journal.

Once more the starting date of demand is all important but to maintain the idea of costing only instantaneous demand changes, the definition of marginal cost amounts to calculating the difference made to costs by postponing (or advancing) the starting date of a permanent output increment. The same cost elements are involved - capacity and running costs - but now we have to consider only the fraction of the total present worth of costs that is avoided (or incurred) by postponing a starting date (or advancing it).

Looking at the expressions (a) and (b) we used above, try to work out the effect on C* of postponing θ by one period. Beginning with expression (a) (for the case where a new consumer appears in a lull in demand requiring no new capacity to be installed), the only term depending on θ is the present worth of the stream of marginal generating costs. Hence varying θ saves or incurs only an instant's marginal generating cost i.e.

(a) instantaneous dynamic marginal cost is marginal generating cost in period θ.

However if - as in case (b) - the consumer's appearance requires extra capacity to be built, the change in θ will affect the present worth sum of capacity and running costs together:

(b) instantaneous dynamic marginal cost is

(i) marginal generating cost in period θ on the adjusted system

plus (ii) year θ's fraction of capacity costs.

Despite the fact that we now have a different conception of marginal cost, the crucial determinant of its value is still the time period θ in which demand appears. Case (a) would require a lower price than case (b). However the consumer is now only available for one instant for the recovery of costs and cannot be offered a variety of tariff schemes; he can only be offered a different price for each time period in which he might appear.

This then is the economist's marginalist approach to tariff setting and his policy prescription is to announce the tariff as a set of forward prices for the planning horizon so consumers are made aware of the fact that it is the particular time period of their extra demand that determines how costly it is for the electricity authority to provide them with supply for that period.

This particular piece of analysis can be tidied up by examining a simple model of the measurement of instantaneous dynamic marginal cost.

(a) the consumer appears at an off peak period

marginal cost is: marginal generating cost on the existing system without new capacity: α_t

(b) consumer's appearance requires new capacity because he appears at a peak

marginal cost is: marginal cost of building and running new capacity this period: β_t

β_t has two elements:

(i) a running cost element which is lower than α_t for case (a) because the adjusted system takes advantage of the "up-to-dateness" of the new plant.

(ii) an amortization element which in future years will fall because the continuation of technological progress requires that the largest amount of recovery of capital costs be in the first year of operation of the new capacity.

Now suppose we have already reached an optimum of C* and are looking at infitesimally small changes in demand around the optimum. If our optimal investment policy has properly allowed for the appearance of infitesimally small changes in demand then it turns out that $\alpha_t = \beta_t$. In other words, the cost of working old capacity a little harder for an instant is very nearly identical to the cost of installing new capacity and running it for that instant after first instant's amortization charge has been added on to the new capacity's relatively lower running cost. (The precise mathematics of this is indicated in Part II of my thesis.)

The models that have appeared in the electricity pricing literature contain the germ of the above ideas though they are usually expressed in terms of the static diagrams of traditional microeconomic theory. The basic idea of a single price for electricity for a given period making up a tariff structure of hourly or periodic prices still emerges. To set the result in the context of that literature I now present a brief summary of the more well known contributions to peak load pricing theory by economists.

PEAK LOAD PRICING MODELS IN THE ECONOMICS LITERATURE

The peak load pricing problem has been the most famous of the electricity tariff problems tackled by professional economists and the "solution" has been suggested several times, one of the best treatments being the early work of Lewis (1941). The models examine electricity demand in several subperiods but the chief unit of time is a cycle of subperiods. Each subperiod has its own demand curve

$$p_j = p_j(x_j) , \quad j = 1 \ldots n$$

Capacity need only satisfy the largest of these demands, which means the "off peak" demands are not "responsible" for some part of the capacity of the system, and in "off peak" subperiods some of this total capacity is unused. The problem is how to decide on optimal capacity and then allocate its costs.

One solution might appear to be to allocate all capacity costs to one demand and only operating costs to other demands in off peak periods. The tariff implication of this is two variable charges per KWh

$$p_1 = c_1 + \beta$$

$$p_2 = c_2$$

p_1 bears the operating costs of subperiod 1, c_1, plus all capacity costs β whereas p_2 bears only the operating costs of subperiod 2, c_2. This is in fact not the general solution, but only a solution to a particular case. However Houthakker (1951) believing that a practical constraint was the necessity to have no more than two running charges, argued for this solution. Several subperiods might pay p_1 and several p_2 but no other prices would be charged. <u>This means that several different loads may bear the same price</u>. Let us call this the Houthakker solution.

To obtain the so called "correct" solution which has the <u>property that the same load may be charged several different prices</u> we can use the model of Littlechild (1970).

Once again the partial equilibrium welfare criterion is to maximize willingness to pay less social cost. For one subperiod we have:

$$B = \int_{o}^{x_j} p_j (x_j)\, dx_j - c_j x_j$$

where c_j is operating cost per unit of output. But how do we decide optimal capacity, which will be equal to at least one of the demands in the subperiods. Steiner's solution to this was to measure the "demand for capacity" in the manner of the demand for a public good; by vertical summation of the demand curves (Steiner 1957). He therefore obtains a measure of total social benefit from a given amount of capacity, and this capacity is available to satisfy any subperiod demand: it exactly provides for peak periods' demand and more than provides for off peak periods' demands. This at least illustrates the nature of the final solution, although there may be problems in actually attaining it. Total benefit in other words is

$$\sum_{j=1}^{j=n} \int_{o}^{x_j} p_j (x_j)\, dx_j \qquad \text{where } j \text{ is an index of subperiods}$$

We still have to consider capacity cost. The easiest way is to think of this as an annual equivalent charge. Think of a piece of new equipment costing K dollars. K can be thought of as the present value of a constant time stream of payments:

$$K = \beta \frac{(1+i)^n - 1}{i\,(1+i)^n}$$

where i is the rate of discount and n is the life of the equipment.

Hence

$$\beta = \text{annual (or per time unit) cost of capacity} = K \frac{i(1+i)^n}{(1+i)^{n-1}} ;$$

as $n \to \infty$, $\beta \to iK$ and if we also allow for depreciation in physical terms, as $n \to \infty$, $\beta \to (i+d)K$ where d is the physical rate of decay of the equipment.

Let us call the amount of capacity purchased y and make explicit the constraint that output may never exceed capacity:

$$x_j \leqslant y \quad , \quad j = 1 \ldots n$$

Hence our problem can be written as the non-linear program

maximize $L(x_1 \ldots x_n, y, v_1 \ldots v_n) =$

$$\sum_{j=1}^{j=n} \int_o^{x_j} p_j\,(x_j)\,dx_j \;-\; \sum_{j=1}^{j=n} c_j x_j \;-\; \beta y \;+\; \sum_{j=1}^{j=n} v_j\,(y - x_j)$$

for x_j , $y \geqslant o$, $j=1 \ldots n$

Now writing down the Kuhn-Tucker conditions for the local maximum to this problem (for an explanation of the Kuhn-Tucker conditions see the Appendix to Part II).

$$\frac{\partial L}{\partial x_j} = p_j\,(x_j) - c_j - v_j \leqslant 0$$

for $j = 1 \ldots n$

$$x_j \frac{\partial L}{\partial x_j} = x_j\,(p_j\,(x_j) - c_j - v_j) = 0$$

$$x_j \geqslant 0$$

$$\frac{\partial L}{\partial y} = -\beta + \sum_{j=1}^{j=n} v_j \leqslant 0$$

$$y\frac{\partial L}{\partial y} = y(-\beta + \sum_{j=1}^{j=n} v_j) = 0$$

$$y \geqslant 0$$

$$\frac{\partial L}{\partial v_j} = y - x_j \geqslant 0$$

$$v_j \frac{\partial L}{\partial v_j} = v_j(y - x_j) = 0 \qquad \text{for } j = 1 \ldots n$$

we can simplify these by assuming in turn

(i) $x_j > o$

$$\text{then } p_j = c_j + v_j$$

(ii) $y = x_j$

$$\text{then } v_j > o$$

(iii) $y > x_j$

$$\text{then } v_j = o$$

(iv) $y > o$

$$\text{then } \beta = \sum_{j=1}^{j=n} v_j$$

These then reduce to the following simple rules: from (i) and (ii) if output is equal to capacity, price exceeds operating cost by some element v_j which will restrict output to capacity. In other cases, prices equals operating cost only. From (iii) and (iv) the sum of all these market clearing excesses of price over operating cost should just cover the cost of new capacity. The welfare interpretation of the v_j is

$$\frac{\partial L}{\partial y}$$

the additional social benefit of a unit of new capacity, as measured by the intensity of its total evaluation among consumers.

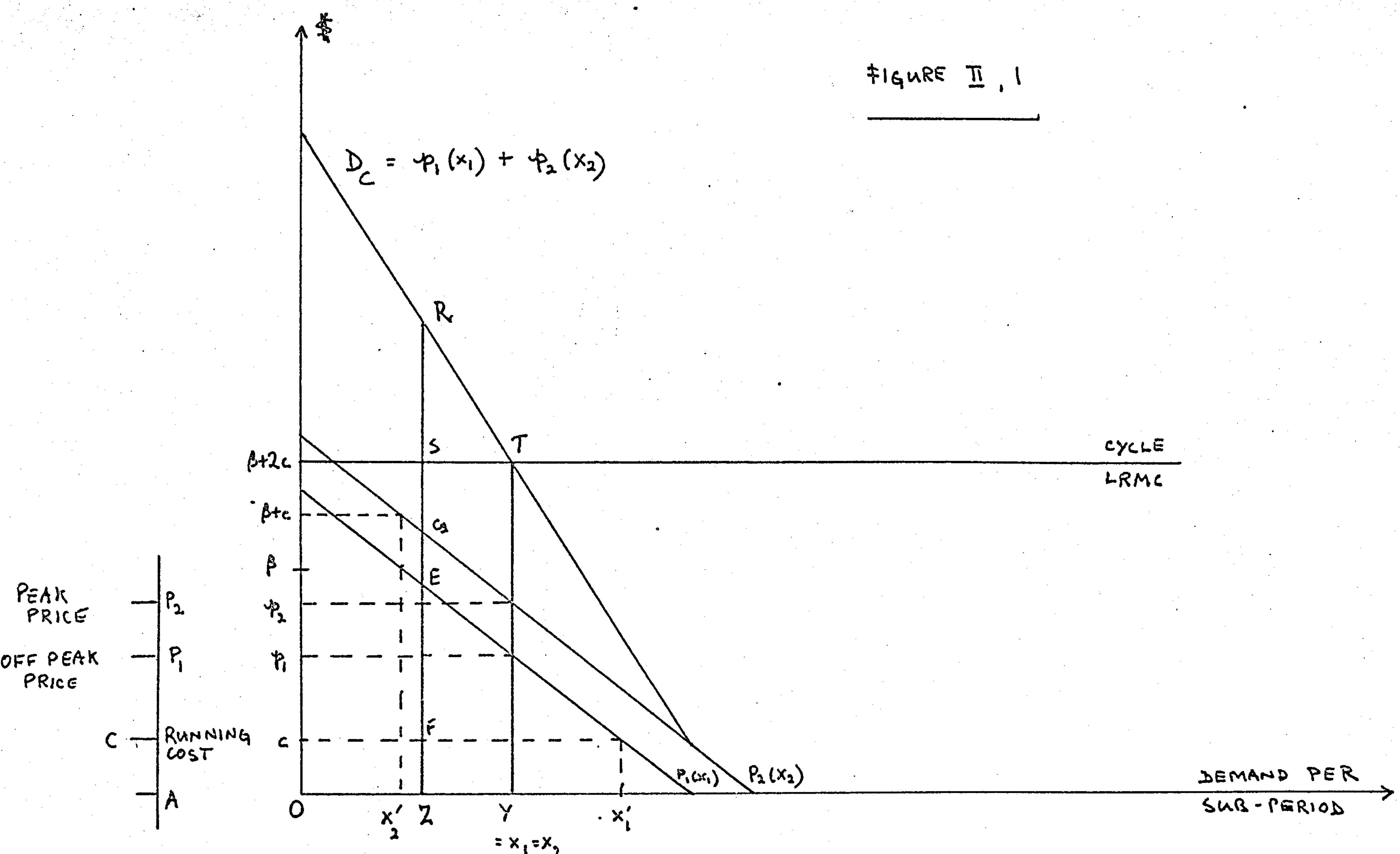
This is the basic solution arrived at several times in the literature. The most convenient diagrammatic formulation is that of Heishleifer (1958) - see fig. II, 1, which has now become quite famous. For simplicity operating costs are assumed to be the same and constant in both subperiods: Oc, and the two (these could be n) subperiod demand curves are p_1 (x_1) and p_2 (x_2). The demand for capacity curve is $D_c = p_1 (x_1) + p_2 (x_2)$, the vertical summation of the demand curves. In terms of the classical exposition, capacity costs are joint to both periods and we can draw up a "joint" LMC curve which in this simplified case is $\beta + 2c$ - cycle LMC (note the horizontal axis measures output per subperiod).

Suppose for the moment capacity is OZ. The market clearing prices are ZG and ZE and demand is restricted to capacity in both subperiods. The distances EF and GF are the illustration of the Kuhn-Tucker multipliers v_1 and v_2. Since the distance (ZE + ZG) = ZR exceeds $\beta + 2c$ = LMC, the Kuhn-Tucker criterion

$$\beta = \sum_j v_j$$

is not satisfied and capacity may be increased; the extra social benefit is illustrated by the triangle RST. Capacity is optimal when no additional consumers' surplus can be obtained, i.e. at OY. At OY capacity, output is again the same in both subperiods $OX_1 = OX_2 = OY$ and prices charged are AP_1 and AP_2 (see side of figure); both exceed operating costs, but by different amounts since the marginal evaluation of capacity in subperiod 2 exceeds that of subperiod 1.

FIGURE II, 1


$
$D_C = p_1(x_1) + p_2(x_2)$
R
S
T
$\beta+2c$
CYCLE
LRMC
$\beta+c$
G
β
E
PEAK PRICE
P_2
p_2
OFF PEAK PRICE
P_1
p_1
F
C
RUNNING COST
c
A
$P_1(x_1)$
$P_2(x_2)$
DEMAND PER SUB-PERIOD
O
x_2'
Z
Y
$= x_1 = x_2$
x_1'

Clearly then the solution does not necessarily involve the recovery of capacity costs only from the peak demand, charging off peak demand only for operating costs. This could lead to severe difficulties. Consider this Houthakker solution in fig. II,1. It would appear that $p_1(x_1)$ is charged a price of 0^c and $p_2(x_2)$ is charged $\beta + c$. This leads then to outputs $0x_1'$ and $0x_2'$; the "peak" has now shifted and what was previously off peak demand now exceeds capacity and previous 'peak' demand now falls short of capacity.

Lewis (1941) seems to have been the first to arrive at the 'correct' solution. Houthakker developed his solution - which is correct when $p_1(x_1)$ falls short of capacity even when charged no capacity costs - in 1951. In 1949 Boiteaux had obtained the general solution for a cost minimizing objective function and later Steiner, Hirshleifer and Williamson all reached the same point, Williamson's being perhaps the most general solution.

The tariff implication of this recommendation is a vector of variable charges per Kwh

$$(p_1 \cdots p_j \cdots p_n)$$

where j is an index of the subperiods of a demand cycle; the larger is demand in a subperiod, the more are capacity costs charged to that subperiod's demand. A current illustration is the Victorian time of day tariff:

7 a.m. - 11 p.m. (first 200 Kwh per month)	4.85 ¢ Kwh
11 p.m. - 7 a.m. (all consumption)	1.00 ¢ Kwh

Especially in the European literature there has long been a preoccupation with another objective of tariff setting - recovering a deficit from marginal cost pricing arising from increasing returns to scale.

Henderson (1947) seems to have been one of the first to tackle this problem which arises when LAC > LMC.

His general solution was to have

average price > marginal price; a typical example of this is the <u>two part tariff</u>. If you like, this is really the "bridge building solution". Charge only marginal cost for a service and provide the service in the first place if total consumers surplus exceeds total cost of building the service. Recover this surplus by charging a fixed sum to all potential users who may then use the facility by paying a variable charge equal to marginal cost.

The tariff implication is

fixed charge: A

variable charge: p per Kwh

as an example we have: (Tasmania)

supply charge per room per quarter	41.00 ¢
charge per unit (first 300 units)	2.41 ¢ per Kwh

The fixed charge finances the service.

This is quite simple. However, the two part tariff in electricity supply is a great deal older than Henderson's paper. It was introduced in the 19th century by Hopkinson as a "scientific" way of allocating costs and solving the peak load problem.

Consider Houthakker's solution to the peak load problem. The two part tariff was an attempt to achieve this by charging one running rate (¢ per Kwh) and a fixed charge "proportionate to the greatest rate of supply the consumer will ever take". The essentially fallacious point is associating peak demand with certain consumers rather than recognizing that all consumers may be involved in both peak and off peak demand. There is no necessary correlation between the greatest supply a consumer will take and peak demand on the generating system.

Despite its illogicallity this tariff has dominated electricity supply pricing all over the world since its first introduction at the close of the nineteenth century. All sorts of reasoned arguments are given for two part or multipart charging, and it is generally favoured by engineers. A good example of the arguments advanced in its favour appears in a United Nations publication "Electricity Costs and Tariffs: a general study" (1972) designed to be used by generating authorities in developing countries. The authorities begin by noting that there are chiefly two types of "cost" in electricity supply: fixed and variable corresponding to the two "services" provided: readiness to supply and supply itself. It is reasonable therefore that two types of price be charged - a standing charge and an energy charge with only the latter varying with output taken. The arguments here are entirely different from Henderson's quoted above. Henderson is trying to capture the consumers' surplus at maximum use to cover the monetary costs of a project for which it is already known that social benefit outweighs social cost. The present argument is based on a (mistaken) view of the costs of supplying electricity. The argument is in the language of cost accountancy - essential for managers' monitoring of current performance - but no guide to decision making. The first fallacy consists in differentiating readiness to supply from supply itself. Decisions about policy are concerned with meeting a given energy demand; nothing more. Capacity is varied to meet demand in the cheapest way, hence the notion of fixed costs associated with readiness to supply is purely arbitrary and misleading; it fails to recognize that price policy is designed to signal to the consumer that the authority is carrying out planned changes in capacity according to forecast demand variations. The true electricity supply problem is not to cover costs of fixed capacity but, as we have seen, to accommodate fluctuations in demand when output cannot be stored.

However it may be possible under certain conditions to tackle the peak load pricing problem with a two part tariff (Lewis (1941)). To do this the standing charge would have to be based on the amount the consumer will take at the time of maximum system demand. This could then give the same sort of solution as the firm peak case described above in figure II,1.

This is the tariff advocated by Posner (1973) and used in bulk supply tariffs for both electricity and gas in Britain. In its ideal form the standing charge or charges depend on "nominated" consumption at the next peak of system demand. This does signal peak capacity costs (which is what is required) and carries an incentive for the customer to shift his peak away from the time of system peak. But while it does signal some true cost information, it does not signal as much as the proper time of day tariff. Its proponents argue that it is less complex than a time of day tariff (this is Posner's defence in particular). However the sort of time of day tariff they have in mind is the purist's vast array of forward prices. A more limited set of time of day charges could be no more complex than the sort of multipart tariffs at present in use and could still signal much more information to users.

In any case, the standing charge for Tasmania that we described above is hardly adequate as a system maximum demand signal. Charges related to such parameters as room size or quality of house bear only the most tenuous relationship to a customer's contribution to system maximum demands. Some fixed charges are however related to maximum demand e.g.

for consumer's maximum demand in KW per month, metered on special meter on consumer's premises	\$ 2.70
running charge (first 1000 Kwh per month)	3.51 ¢ per Kwh

This applies to industrial consumers in South Australia. But again, this maximum demend measure is not necessarily related to maximum demand on the system, and its usefulness as a solution to the peak loading problem is probably very limited.

Nevertheless, this concept has been so popular that it turns up in a variety of guises. Other forms include block tariffs, or quantity discounts. A different running charge operates in each block. "Earlier" blocks e.g. first 50 Kwh per month may be charged at a relatively high rate with charges diminishing thereafter. The tariff implication is

For demand = X^* : $p(x_1) \ldots p(x_j) \ldots p(x_n)$

with $\sum_{j=1}^{n} x_j = X^*$ and earlier charges reflect more of capacity costs than later charges.

The Henderson argument of course was for two part tariffs as a solution to the financing problem. However, the two part tariff or block tariff is not the only theoretical solution to the financing problem. Consider an individual consumer's demand curve for electricity (assume no peak/off peak loading problem for convenience)

$$p_k = p_k(x_k) \qquad k = 1 \ldots m$$

where the index k represents particular consumers.

The supplier's objective is again assumed to be the maximization of benefits less cost. In this case benefits are the <u>horizontal</u> summation of the individual demand curves and <u>an additional constraint is the provision of a surplus over costs</u>:

i.e.

maximize $L(x_1 \ldots x_m, a) =$

$$\sum_{k=1}^{k=m} \int_0^{x_k} p_k(x_k^1)\, dx_k^1 - C(x_1 \ldots x_m) + a\left[\sum_{k=1}^{k=m} p_k(x_k) - C(x_1 \ldots x_m) - S\right]$$

where $C(\;)$ is the general social cost function and S is the financial surplus required.

Using the expression:

$$e_k = -\frac{\partial x_k}{\partial p_k}\frac{p_k}{x_k}$$

for the compensated price elasticity of demand, the optimality conditions can be simplified as follows (Baumol and Bradford 1970)

$$\frac{p_k - C_k}{p_k} = \frac{a}{(1-a)\ e_k}$$

where k = 1...m is an index of individual consumers

and where $a = \frac{\partial L}{\partial S}$, the marginal social utility of a \$ 1.00 of surplus, and $C_k = \partial C/\partial x_k$.

In other words each consumer is charged a price above marginal cost, the excess varying among different consumers according to each consumer's elasticity of demand. The tariff implication of this method of financing is a vector of running charges

$$(p_1 \cdots p_k \cdots p_m)$$

Where the index describes different consumers or consumer groups.

This again is widely used: e.g. in all states, the marginal running charges vary among consumer classes of the type:

- residental consumers
- commercial consumers
- industrial consumers
- religious organizations
- all consumers together

We can sum up this summary of the theoretically suggested pricing schemes by looking at the tariff implications they derived:

(i) time of day tariffs, with peak demand prices showing a relatively greater reflection of capacity costs

(ii) (a) two part tariffs, with the standing charge reflecting capacity costs

(b) use of maximum demand charges, reflecting capacity costs

(c) use of block tariffs, with "earlier" blocks reflecting capacity costs relatively more

(iii) running charges that discriminate amongst different groups of consumers.

We now want to obtain some empirical measure of the application of these schemes.

CHAPTER III

ELECTRICITY TARIFFS IN PRACTICE

There are several possible ways of obtaining a measure of the practice of theoretical principles. I have chosen one which depends on rather heroic assumptions but is nevertheless quite useful.

Part III of this thesis outlines a basically recursive model of electricity demand and supply. The second equation block of the simple model represents a purely cost based set of tariffs. Now this pricing hypothesis has found widespread use in recent years in macroeconometric model building where prices are usually related to a measure of normal unit labour costs and normal unit capital costs (e.g. Norton and Others, 1970; Helliwell et.al. 1969). More generally I have assumed tariffs are related to planned unit costs in roughly the categories:

- fuel for generation
- 'other' generating costs; chiefly labour costs
- capacity costs

Following the basic macro-modelling approach:

$$p_t = \alpha + \sum_i \beta_i x_{it} + v_t$$

where p_t is price and x_i is planned unit cost of the i^{th} factor of production, and v_t is a disturbance term.

Using the data of the Electricity Supply Association of Australia the basic categories of unit cost corresponding to our theoretical analysis that can be singled out are:

UFCP - planned unit fuel for generation costs

UOCP - planned unit other costs

UICP - planned unit interest costs

UDCP - planned unit depreciation costs

Clearly (UICP + UDCP) correspond to "capacity costs" of the theory (see remarks on β CH.II) and (UFCP + UOCP) correspond to operating costs.

I would like to postpone for the moment the actual measurement of planned costs, given a knowledge of actual costs; I will explain this after considering the price cost relationships.

There are several reasons why the normal cost pricing hypothesis has been used in macro econometric modelling: the basic hypothesis is:

"... prices move with long run costs and do not change because of variations in demand or cost which are thought to be temporary ..."

(Godley and Nordhaus 1972)

In fact there have been many arguments for using only infrequent variations in electricity tariffs because there are administrative and other economic costs involved in changing tariffs and adapting to them. Consequently the 'normal' cost hypothesis appears quite attractive as an explanation of the relatively slow movements in administered electricity tariffs. An examination of the Australian data suggests that on average, some major alteration in tariff schedules occurs every three years.

While adopting the hypothesis, I have tried to develop what I believe is a better measure of normal costs - I have called it planned costs, and I will set this out after considering the tariff relationships. The reason for this ordering is that the method of measurement does itself cast light on some problems raised in the application of theoretical tariffs. First I would like to consider the price data.

TARIFF DATA

The data consist of numbers of different tariff series charged to various categories of consumers in most parts of Australia over the last twenty years. Unfortunately every state could not be represented. No published tariff series covering Queensland was obtainable at the time of writing. In New South Wales there are so many distribution authorities (though some are more important than others) that it was not possible to relate published tariffs to the existing State wide cost data. This leaves four states in the sample: Victoria, South Australia, West Australia and Tasmania, and some very good 'hard' data on electricity prices was obtainable for these States covering the period of the 1950's and 1960's. From this data, it proved possible to set up electricity price series of much more use and complexity than conventional average revenue figures, and this is one of the most interesting aspects of the study. All in all, nearly 300 different rates for a unit of electricity have been used in these 4 states over the post war period, an astonishing degree of price discrimination which bears little relationship to the actual marginal costs of the units produced. In many cases reported tariffs overlapped for different classes of consumers and often a tariff was used for a few years, discontinued for a period and then perhaps used again in a slightly different category. Very often the amounts at which quantity discounts became available were altered. Nevertheless a substantial number of complete continuous, usable series have been obtained, pertaining to different consumer categories, seasons and times. These form the basis of the present study. The coverage of the tariffs will become plain in the reported results, where the following comprehensive coding system has been adopted.

STATES: There are 4 states covered

VICTORIA	:	V
SOUTH AUSTRALIA	:	S
WEST AUSTRALIA	:	W
TASMANIA	:	T

CATEGORIES: There are 5 distinct categories of consumers used

RESIDENTIAL	A
INDUSTRIAL	B
COMMERCIAL	C
COMMUNITY SERVICE, RURAL, OR OTHER	D
ALL CONSUMERS	E

QUANTITIES: Quantities for bulk discounts are in ascending order, beginning with the most expensive: 1, 2, 3 . . .

PREFIX: One prefix, R, is used to denote a tariff only available in restricted hours, during the day or night.

RATIO: The ratio of an off peak, or night time restricted hour tariff to its peak or day time alternative is denoted: PI

The coding is then in the following order:
PREFIX (Restricted hour: CATEGORY: STATE: QUANTITY if used)
so that for example:

RAV1 is the most expensive quantity block (1) of a Victorian (V) restricted hour tariff (R) applicable to residential consumers (A).

CBT3 is the third (3) quantity block of a Tasmanian (T) tariff applicable to commercial (C) and industrial (B) consumers.

A full listing of the published name of each tariff used in this study, along with a copy of the data, which, it is felt, will be of interest to other workers in this field, is obtainable from the author.

OVERALL REGRESSION PERFORMANCE

Returning to estimation of tariff-cost relations in practice, it can be recalled that the basic model was

$$p_t = \alpha + \sum_i \beta_i x_{it} + v_t$$

with x_{it} being unit factor cost.

Despite some evidence of multicollinearity (it was decided not to drop "wrongly" signed variables to guard against specification error) the overall performance was fair. Usually about 80-90% of the variance in tariffs was explained by unit cost measures. The regressions however were only part of the task. Regression coefficients are not enough to indicate the relative strengths of relationships with different variables; some systematic measure of individual contribution is required. Several statistics are available (Goldberger 1964 pages 197-200) and the one chosen was the partial correlation coefficient:

$$r_{p\,x_j}$$

the square root of the proportion of variation in p (after all other regressors have been allowed for) which was accounted for by x_j. This statistic is preferred by both Goldberger (1964) and Theil (1971) to other forms of decomposition of the total explanation. It is equivalent to the ordinary correlation coefficient between x_{jt} and the residuals:

$$(p_t - \sum_{i \neq j} \beta_i x_{it} - \alpha).$$

These experiments can be summed up as follows:

(a) a detailed breakdown of tariff types was obtained.

(b) each type was regressed against measures of planned unit fuel, "other", interest and depreciation costs.

(c) partial correlation coefficients for those cost variables entering each equation with a positive sign were calculated and these coefficients were used to examine the testable hypotheses of the theoretical treatments of electricity prices.

In the analysis which follows, especial attention is paid to those variables for which the partial correlation coefficient with a tariff category is at least as large as 0.300. This figure corresponds, for the sample size used here, to a t - test critical rejection level of the hypothesis that the regression coefficient does not differ from zero of about 25%.

It is easy to argue that this approach is rather crude, but given the availability of data it can be expected to yield some helpful insights into tariff setting behaviour. The full results are analysed by cost and tariff categories for each state in tables III, 1, a, b, c and d (Victoria, South Australia, West Australia and Tasmania).

Table III, 1 (a)

TARIFF NAME	PARTIAL CORRELATION OF TARIFF AND PLANNED COST UFCP	UOCP	UICP	UDCP	R^2 OF THE ESTIMATING EQUATION
1. AV1		0.584		0.008	0.9029
2. AV2		0.466		0.015	0.9517
3. AV4		0.539			0.9612
4. BV3		0.270		0.182	0.9258
5. BV6	0.386	0.150		0.445	0.7245
6. BV8 (MDCH)		0.139	0.001		0.2357
7. RCV10	0.339	0.005	0.194	0.302	0.8778
8. CV1	0.413		0.224	0.287	0.8179
9. CV2		0.284		0.169	0.9273
10. CV6		0.247		0.182	0.9129
11. CBV1		0.269		0.184	0.9259
12. CBV2		0.297		0.147	0.9252
13. CBV3		0.324		0.175	0.9354
14. CBV4	0.060	0.157	0.053	0.220	0.9036
15. REV1	0.597		0.168	0.474	0.5672

Table III, I (b)

TARIFF NAME	PARTIAL CORRELATION OF TARIFF AND PLANNED COST UFCP	UOCP	UICP	UDCP	R^2 OF THE ESTIMATING EQUATION
1. AS2	0.019	0.509		0.347	0.6526
2. AS3	0.096	0.366		0.438	0.7038
3. AS4	0.428	0.443	0.069	0.343	0.7920
4. AS5	0.204	0.589		0.397	0.6257
5. RBS1		0.394		0.120	0.4547
6. RBS2		0.578			0.4214
7. RBS3	0.411		0.321	0.077	0.7405
8. RBS4	0.398		0.426	0.002	0.7866
9. ES1	0.194	0.500		0.459	0.6270
10. ES2	0.072	0.378		0.429	0.7175
11. ES3	0.082	0.370		0.418	0.6909
12. ES4	0.112	0.611		0.378	0.6489
13. ES6	0.173	0.520		0.449	0.6381
14. ES7		0.422		0.343	0.7423
15. ES8	0.068	0.578		0.368	0.6408
16. RES9		0.434		0.341	0.7703
17. RES10		0.416		0.354	0.7332
18. RES11	0.063	0.539		0.196	0.4240

Table III, 1 (c)

TARIFF NAME	PARTIAL CORRELATION OF TARIFF AND PLANNED COST				R^2 OF THE ESTIMATING EQUATION
	UFCP	UOCP	UICP	UDCP	
1. AW1	0.888	0.385	0.326		0.9307
2. BW1	0.950		0.651		0.9567
3. BW2	0.683	0.365	0.132		0.8072
4. BW3	0.855	0.186	0.093	0.185	0.8944
5. CBW1	0.870		0.600		0.8546
6. EW1	0.785	0.331	0.088	0.073	0.8650
7. EW3	0.910		0.041	0.432	0.9088
8. EW4	0.689	0.399	0.097		0.8234

Table III, 1 (d)

TARIFF NAME	PARTIAL CORRELATION OF TARIFF AND PLANNED COST				R^2 OF THE ESTIMATING EQUATION
	UFCP	UOCP	UICP	UDCP	
1. AT2		0.382	0.052		0.8988
2. BT3		0.034	0.362		0.8794
3. BT4			0.432		0.8026
4. CBT1		0.095	0.358		0.9031
5. CBT2		0.109	0.321		0.8917
6. CBT3			0.387		0.8422
7. DT1			0.397		0.8450
8. DT2		0.001	0.376		0.8729
9. RET2		0.035	0.366		0.8838
10. RET4		0.298	0.106		0.8843

Before examining the details of the different categories, there is one result that we can expect to show up straight away. It has been a frequently reported characteristic of the post war history of the electricity authorities that it is inflation that most often prompts an authority to restructure its tariffs. For example, both South Australia and Victoria have recently (1973) announced tariff changes based on labour cost escalations. Recalling that labour costs are measured in the "other" costs category, it is to be expected on the basis of the authorities' public statements that the cost category UOCP (planned unit "other" costs) might have greatest influence on tariff movements. Table III,2 shows the overall picture but table III,3 shows that there may be considerable interstate differences in the impact of cost inflation on tariffs.

However the important question to ask is whether - given that tariffs have risen to recover cost rises due to inflation - these changes have overwhelmed the application of the theoretically suggested tariff schemes.. What has inflation recovery done to the principles (economic based or engineering based) of tariff determination?

It can be shown that when there are theoretically determined cost-tariff ratios the recovery of inflation in costs need not distort the underlying relationships. The appendix to this part of the thesis investigates the effects of cost inflation on some aspects of tariff setting. The results are quite simple to summarize:

Table III,2

Partial Correlation with different tariff types:	Cost Categories UFCP	UOCP	UICP	HDCP
(a) No. of positive coefficients	27	41	27	33
(b) As % of 128	21·	32	21	26

Table III,3

POSITIVE CORRELATION WITH TARIFF	UFCP	UOCP	UICP	UDCP
VICTORIA (a) No. of Positive Correlation Coefficients	5	13	5	12
(b) As % of 35	14	37	14	34
SOUTH AUSTRALIA (a) No. of Positive Correlation Coefficients	13	15	3	17
(b) As % of 47	28	32	6	36
WEST AUSTRALIA (a) No. of Positive Correlation Coefficients	9	6	9	4
(b) As % of 28	32	21	32	14
TASMANIA (a) No. of Positive Correlation Coefficients	(n.a.)	7	10	0
(b) As % of 17	(n.a.)	41	59	0

(i) inflation at equal rates in operating and capacity cost categories should leave tariffs relatively in same relationship with each other.

(ii) faster rates of inflation in particular cost categories should be reflected most in the tariffs with highest relative weighting of those costs..

Thus inflation in costs <u>can</u> be recovered without distorting the basic principles of tariff setting.

Nevertheless this theoretical result may not be practised by electricity authorities who may - in trying to cover inflation - pay scant attention to the principles of tariff setting.

DETAILED ANALYSIS OF THE RESULTS

Begin with the tariff principles derived from the peak loading problem; as we have seen there are several ways of treating the complication of varying loads.

Most authorities do consider the possibility of off peak tariffs. West Australia however is an exception. There are no off peak options whatsoever available to any West Australian consumers. The other states sampled use time of day tariffs to a varying degree for different consumers. None however use more than 2 separate time periods; usually these are

7 a.m. - 11 p.m. : Peak (day)

11 p.m. - 7 a.m. : Off Peak (Night)

This suggests they adopt at most the "wrong" Houthakker solution to the problem and consequently charge the same price for different amounts of supply.

The off peak rates sampled were the following:

VICTORIA:

RCVIO	Commercial and industrial night rate
REVI	All customers night rate water heating

From 1967 Victoria introduced a number of new off peak options (see below).

SOUTH AUSTRALIA:

RBS2) RBS4)	industrial night rate (block discount categories)
RES9) RES10)	all customers night rate (block discount categories)
RES11	night rate water heating

TASMANIA:

RET2	two rate supply, all customers: night rate

Table III,4 summarizes the results on these tariffs. Tasmania is the clearest example of a misuse of night rate tariffs. Capital costs (interest) have had much greater impact than operating (other) costs. Comparing table III,4 with the overall results of table III,1, the Tasmanian night rate has reacted as much to capital cost changes as

TABLE III,4 Off Peak Tariffs: partial correlation coefficients with different categories of unit factor cost.

TARIFF	UNIT FACTOR COST UFCP (fuel)	UOCP (other)	UICP (interest)	UDCP (depreciation)
RCV10	0.339	0.005	0.194	0.302
REV1	0.597	-	0.168	0.474
RBS2	-	0.578	-	-
RBS4	0.398	-	0.426	0.002
RES9	-	0.434	-	0.341
RES10	-	0.416	-	0.354
RES11	0.063	0.539	-	0.196
RET2	-	0.035	0.366	-

any of the day rates and less to operating costs changes than several of the day rates, whereas we know from the two period (Houthakker) theoretical solution that off peak rates should have a relatively lower (or zero) weighting of capital costs and possibly a relatively higher weighting of operating costs.

The other states do not err so badly. Both South Australia and Victoria show their off peak rates reflecting operating costs relatively highly but there is still a high reflection of capacity costs. In the second row of table III,4

$$r\ (\text{REV1, UDCP}) = 0.474$$

and from table III,1 this is seen to be the highest of the UDCP correlations. In other words, the Victorian night rate for water heating has shown more influence from capacity (depreciation) costs than any of the other sampled Victorian tariffs. The RCV10 tariff is not much more satisfactory in this respect. RCV10, in its turn, shows the second highest sampled correlation with interest costs of all Victorian tariffs.

For South Australia, there appears to be a stronger relationship between operating costs and off peak tariffs than between capacity costs and off peak tariffs; even so, the relationship with depreciation costs are still relatively high and RBS4 is one of only three South Australian tariffs sampled that shows a correlation with interest costs. If we compare the results of table III,4 with the comments above on the effect of cost inflation, it is apparent that for all three states in question, all tariffs have been made to reflect capacity cost movements relatively less in off peak than in peak tariffs, as should happen if the theoretical recommendation that capacity costs have a low or zero weighting in off peak tariffs is to be followed. The chief cost escalations have been in:

VICTORIA:	"other" and depreciation costs
SOUTH AUSTRALIA:	"other" and depreciation costs
TASMANIA:	interest costs

(see table III,3)

and these have been passed on just as much to off peak as to peak tariffs.

Further on in this part of the thesis some other experiments are attempted to see if more light can be thrown on the peak load pricing problem. Meanwhile some of the other theoretical tariff implications can be examined.

The popularity in utility prices of block tariffs or two part tariffs with or without a "maximum demand" aspect has already been noted. The theoretical implication is that the early blocks or the fixed charge or the maximum demand charge recovers capacity costs relatively more. Most of the tariffs sampled were block running charges, however two of the Victorian tariffs are attempts to recoup capacity costs directly.

(i) AV1 = the residential standing charge per month calculated on the householder's number of rooms.

(ii) BV8 = the industrial maximum demand charge calculated by special metering of the customer's maximum demand.

The partial correlation results extracted from table III,1 are:

	UFCP	UOCP	UICP	UDCP
AV1	-	0.584	-	0.008
BV8	-	0.139	0.001	-

These tariffs seem, on this evidence, to be appallingly misdirected. Both show highest positive relationship with the chief inflationary element in Victorian costs: "other costs" (most labour costs) and little or no relationship with interest or depreciation costs both of which (especially depreciation costs) were rising strongly over the period for Victoria and which have appeared as influences on other tariffs - including off peak tariffs.

We can attempt the same sort of test with different block discount categories - earlier blocks may show more relationship with interest and depreciation costs. Some of the results for the sampled runs of block discounts are shown in table III,5. Tasmania runs contrary

TABLE III,5 Correlations between cost categories and block discount categories for selected tariffs.

Tariff	UFCP (fuel)	UOCP (other)	UICP (interest)	UDCP (depreciation)
	Cost Categories			
VICTORIA				
CV1	0.413	-	0.224	0.287
CV2	-	0.284	-	0.169
CV6	-	0.247	-	0.182
CBV1	-	0.269	-	0.184
CBV2	-	0.297	-	0.147
CBV3	-	0.324	-	0.175
SOUTH AUSTRALIA				
ES1	0.194	0.500	-	0.459
ES2	0.072	0.378	-	0.429
ES3	0.082	0.370	-	0.418
WEST AUSTRALIA				
BW1	0.950	-	0.651	-
BW2	0.683	0.365	0.132	-
TASMANIA				
BT3	-	0.034	0.362	-
BT4	-	-	0.432	-

to the theoretical result, the other states either are not conclusive or just barely exhibit a weighting of capacity costs in the earlier blocks. Again inflation seems to have been recouped from all tariffs regardless of the theoretical weightings suggested

DISCRIMINATION

Finally we can attempt to discover whether these tariff correlations indicate any discrimination among consumer groups in the ways in which costs are passed on.

Empirically we have to distinguish between:

Residential

Industrial

and Commercial

consumers

and in terms of the data we have we are able to provide the following characterisation of any tariff:

(i) its applicability to a certain group

(ii) its correlation with a given cost type.

The test proceeded as follows: for each cost category (fuel, other, interest and depreciation) we have a set of correlation coefficients each with a particular tariff which can be ranked in order of magnitude. But each tariff can also be ranked in the simple measure of its applicability to a standard group of consumers - e.g. residential consumers. Thus tariffs labelled A apply to residential consumers only, tariffs labelled E apply to residential and other consumers while tariffs labelled B and C do not apply to residential consumers at all.

We can compare rankings to decide whether or not cost rises for a particular cost category (e.g. fuel) are associated with tariffs applying to a particular consumer group (e.g. residential consumers). To be quite clear a hypothetical example could be illustrated as follows:

Cost category: fuel costs:

Ranking in order of size of correlation with different tariff types:	Ranking in order of applicability to residential consumers
1. tariffs with label A	1. label A
2. tariffs with label E	2. label E
3. tariffs with label C or B	3. label B or C

thus in the above example there is a perfect correlation suggesting fuel costs are mostly passed on to residential consumers.

The rankings can be compared on the basis of Spearman's rank correlation coefficient and a simple transformation of this tested on the student's t distribution (see Kendall, 1948).

The test applies to results from all states together and picking out those results for which the rank correlation coefficient differs from zero at the 5% level we have as the main result the category of costs which are passed on mostly to residential consumers is "other" costs; whereas interest and depreciation cost rises are mostly passed on to industrial and commercial consumers.

Table III,6 gives some of the main results.

It appears from this fairly crude categorisation that inflation in 'other' costs (Labour and Working Costs) is generally passed on to the residential consumer while cost rises under the interest and depreciation categories are more likely to be passed on to industrial consumers.

TABLE III,6 Coefficients of rank correlation for tariff categories and cost influences.

Cost categories are ranked in order of their correlation with tariff types.
Tariff categories are ranked according to how applicable they are to residential consumers.

Cost categories	Spearman's rank correlation coefficient
Other	0.448
Interest	-0.657
Depreciation	-0.182

CONCLUSIONS FROM THESE RESULTS

We have therefore something fairly straightforward to say on the basis of these tariff results. The theoretical principles of tariff setting suggested the following hypotheses:

(i) off peak tariffs ought to correlate more with running costs than capacity costs.

(ii) standing charges ought to correlate with capacity costs.

(iii) in block discount tariffs early blocks ought to correlate more with capacity costs than later blocks.

The results from our samples indicate a large number of cases where these theoretical principles are abused or ignored in practice. It would perhaps be a more realistic conclusion to say that in practice cost inflation is simply recouped from all types of tariff on an unsystematic basis. It seems reasonable to conclude that these states are not setting their tariffs on the theoretically advocated bases. The questions that then arise are:

(a) Does it matter in terms of cost savings whether or not theoretical principles are adhered to? In other words, are there significant cost savings to be obtained from load spreading? If there are then this abuse of theoretical tariff principles may be socially costly.

(b) Even if there are cost savings to be obtained from load spreading, are tariff adjustments capable of achieving this load spreading - The following Chapters attempt to answer this question.

CHAPTER IV

ELECTRICITY COSTS AND PEAK SPREADING

The sections above discussed the relationship between electricity tariffs and "planned" costs under different categories. It remains to discuss how "planned" costs were calculated but the important question that the previous sections begged was whether peak spreading actually means anything in terms of real dollars. Is it simply an economist's academic problem or would its solution provide real cost savings? Now it happens that in calculating "planned" costs for the previous analysis I believe that some light could be cast on this very question simply as a result of the methods used in the calculation of planned costs. Thus, while the initiating theme of this section is to describe how "planned" costs could be and were calculated, a consequent (and very important) theme is the additional information we can gain on the "dollar dimension" of the peak loading problem. The first comments are on planned costs and in developing them I will explain how the peak spreading aspect is brought in.

Turning now to how I measured planned costs we can consider what the term appears to mean.

The definition of the normal cost mechanism given by Nordhaus and Godley is still sufficiently vague to allow for a great many different interpretations of the measurement of normal costs and their influence on prices. It is soon apparent, on comparing the studies in which the normal cost hypothesis is used, that there is little unanimity on its practical operation; very generally, however, we may determine some of the reasons why particular ways of measuring normal costs have been adopted in previous studies. All of these are deficient in some way, and we are compelled, before adopting one version of normal cost measurement, to re-think exactly what is meant by the hypothesis.

An initial impetus to the normal cost hypothesis in empirical work arose from a desire to capture in econometric models some of the observed stickiness of many industrial prices. Basically this meant introducing a delayed response into cost adjustments. The pioneering study from an empirical point of view was that of Schultze and Tryon (1965). They hypothesised that changes in factor prices would be immediately incorporated into normal costs but changes in factor productivity would only be incorporated if they persisted over a long period. They measured normal unit costs as total factor payments divided by normal output, where normal output was a distributed lag or moving average of actual output. For instance, normal unit labour costs were measured as

$$ULCN_t = \frac{\text{Compensation per man hour}_t}{(\sum_j a_j \; X_{t-j})}$$

where X_t is output at time t. This model was fairly successful, particularly in the category of regulated industries including electric utilities. This measure deflates total factor payments by a measure of normal output. Helliwell and others (1969) adopted the same device but used a more sophisticated measure of normal output. Normal output was achieved when the economy was on its production function; the production function was synthesised from interrelated factor demand equations. The Cobb Douglas model was used and a log linear equation system was solved to give desired factor quantities in terms of relative prices. These factor demand equations when estimated allowed the derivation of a synthetic production function which in turn provided normal output and hence normal costs. Again this approach was relatively successful, though more with respect to labour than capital costs.

A further aspect of price stickiness led to a different method of estimating normal costs. Initially price stickiness merely reflected assumptions about non-competitive market structure. However in large model building a problem that arises is that since output responds to exogenous changes more quickly than employment, price equations that do not use normal costs will have prices falling because of falling unit costs whenever output rises; the "short term productivity" effect, which does not fit with the observed downward stickiness of prices. This phenomenon led to attempts to measure normal costs in terms of actual costs that have been purged of short term productivity changes.

For example, the typical text book micro model of the form is

$$\max PI = px - wL(x)$$

where PI is profit, p is price, x is output and wL(x) is total labour costs. The equilibrium condition is:

$$p = wL'(x) \simeq w\frac{L}{x}$$

so that

$$dp = w\, d(\frac{L}{x}) + \frac{L}{x}(dw)$$

Here price changes reflect both changes in average productivity and in factor prices. However changes in normal costs (or normal supply price) should only, it is argued, reflect changes in factor prices:

$$dp - wd(\frac{L}{x}) = \frac{L}{x}(dw)$$

which suggests that normal costs be measured as actual costs purged of short term productivity effects.

This has been the most popular conception of normal costs in the literature. It was advocated by Eckstein and Fromm (1968) and is the basis of the price equations in Norton, and others (1970) Australian study. It involves regressing actual costs against only those factors thought to determine long run normal cost changes (for example input prices) and taking the predictions from the regression as the series on normal cost. The immediate objection is that is is very arbitrary in its exclusion of possible cost influences. For instance it makes no allowance for changes in costs due to planned movements along a long run scale curve. Such factors will be quite important at the microeconomic level even if they can be regarded as negligible at the aggregate level. Nevertheless the basic assumption that normal costs do not include those short run random disturbances in actual cost regressions is quite attractive. The essential rationale of this method however can be given a more definite if very simple support if it is thought of in terms of the stochastic regressor model of classical least squares. Essentially we are defining normal costs as what actual costs would be without short run, unexpected changes. We have already derived an expression for actual costs as a linear relationship such as:

$$c = a z$$

where z may be a vector of stochastic variables. This relationship is subject to a set of short run changes represented by a random disturbance term u for which we may make the assumptions:

$$E(u / z) = Eu = o$$

Then defining normal costs as what actual costs would be if only the expected influences on costs are considered, we have

$$Ec = E(az + u/z) = az + E(u/z) = az \qquad \ldots\ldots\ldots (1)$$

Hence this definition of normal costs - expectations about actual costs - suggests that the basic rationale of using the predictions from cost regressions is reasonable, but those regressions ought not to exclude arbitrarily any influences on costs that could be planned for. Such influences clearly include planned expansion along the scale curve. In other words, the only legitimate way to use normal costs is to regard them as what actual costs would be without the unplanned, unexplained residual.

On the basis of this rationalization, the series we may adopt for normal costs are the predictions from the regressions of actual costs against all those factors which appear to be important, non random, influences on actual costs: i.e. the predictions from the cost regressions described above. These then are taken as our normal cost data and used to explain movements in reported tariffs.

This therefore describes the method used to calculate planned unit factor cost from actual unit factor costs.

I would now like to digress from considering tariffs to discuss these cost experiments in greater detail.

We take uc(i)p as the predictions from a regression of uc(i) on those factors thought to influence uc(i).

These regressions themselves may therefore throw some light on cost problems in electricity supply. I believe they can be quite useful in discussing the peak load problem. Let us therefore consider these cost investigations.

COST INVESTIGATIONS

The essential aim of this section is to explain average factor costs in the main categories of factors of production reported in electricity authority accounts. This approach to estimating cost relationships owes more to macroeconometric model building than to the well established microeconomic approach to estimating marginal and average cost curves. Basically, we are splitting up the total cost function and ascribing components of total cost to the main factor categories e.g. in terms of capital, labour, and materials.

This gives as a basic relationship for the i^{th} factor of production:

$$uc(i)_t = a_o + a_1 X^*_t + a_2 V^{(i)}_t + u_{it} \quad \ldots\ldots\ldots\ldots \quad (2)$$

where the dependent variable is unit factor cost and the explanatory variables are:

X^*_t : (planned) output which acts as the scale parameter

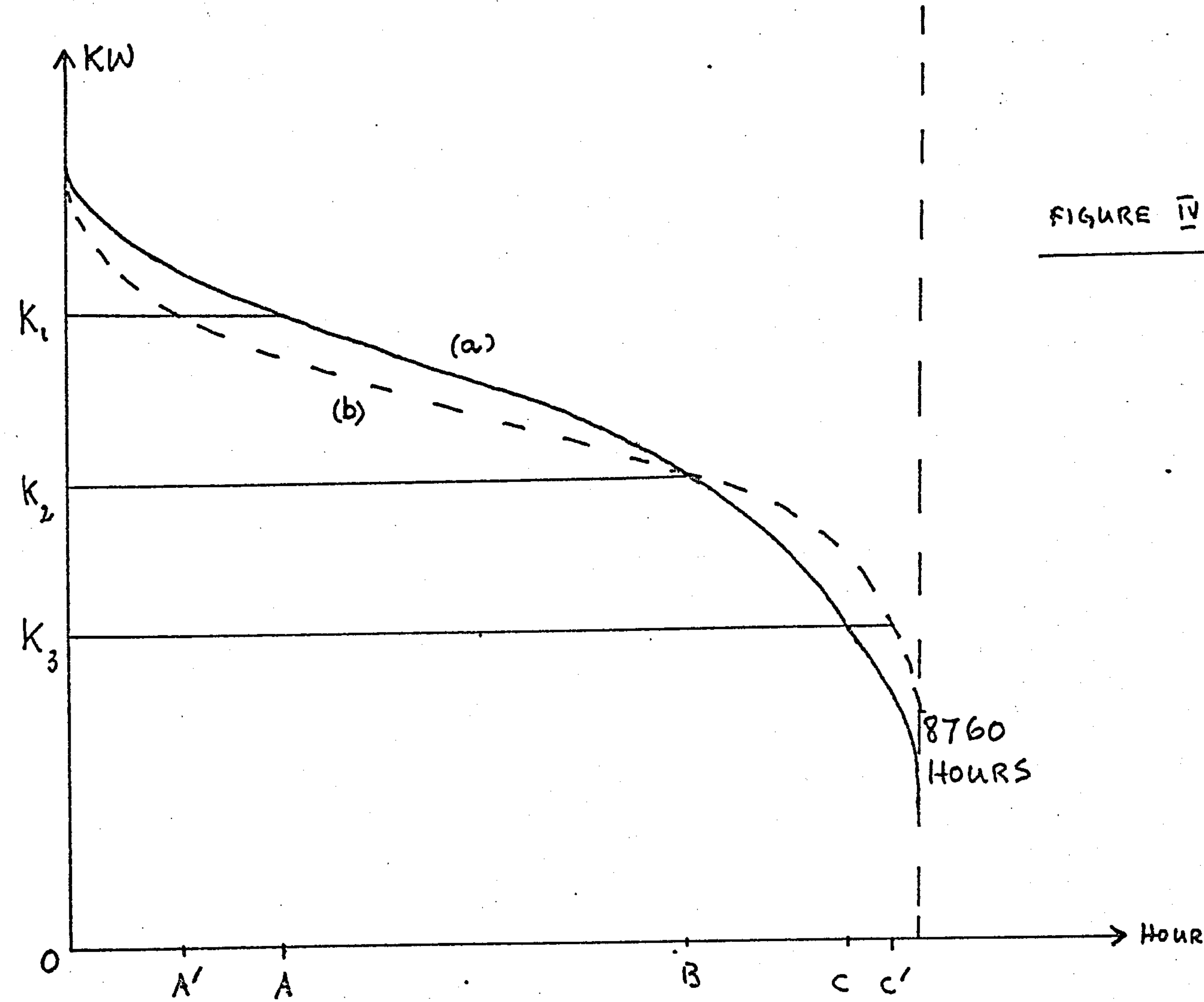
$V^{(i)}_t$: the vector of determinants of the i^{th} factor price

u_{it} : random disturbance term

This is one experimental way of explaining reported cost figures; on the assumption that electricity production reflects a basically fixed coefficient type of technology, it is probably theoretically well based.

This is a standard form of relationship particularly used in macroeconometric studies. However for the case of electricity costs an additional influence needs to be taken into account. This is the ability to reduce costs by spreading peak loads into off peak periods. The relationship in (2) holds for a given degree of capacity utilization but in the sample period of the data on electricity costs this degree of capacity utilization may vary significantly enough to change factor costs substantially. This is best illustrated with the use of a diagram like Fig. IV, 1, which represents a hypothetical

FIGURE IV, 1



HYPOTHETICAL CUMULATIVE LOAD DURATION CURVES cumulative load duration curve. This curve is plotted in capacity-time space. Along the horizontal axis is measured the hours in a production cycle; dealing with annual data we may consider a period of 8760 hours. On the vertical axis is measured the capacity used to satisfy the load on the system. Taking curve (a) to begin with: OK_1 of capacity is in operation for OA hours, OK_2 for OB hours and OK_3 for OC hours, the latter being the largest part of the year. Now, given the "merit order" running schedule of a generating system, base load e.g. OK_3 for OC hours will be met by new base load plant exhibiting running cost savings over older plants. Peak load of OK_1 which lasts for OA hours will be met by both new and old base load and peaking plant (which may be either small high running cost plant e.g. gas turbines, or old conventional thermal plant). Suppose now the load characteristics alter so that curve (b) operates. It reflects, in comparison with (a), a shift from peak to off peak loading. OK_1 capacity is only in use for OA^1 hours and OK_3 capacity rises from OC hours usage to OC^1 hours.

This peak shift may have two effects on costs:

(i) less peaking capacity is used; this means a reduction in capital costs by reduced use of old plant and small peaking plant.

(ii) overall running costs may also fall as more load is met by the newest base load plant, which embodies the greatest cost savings (quasi rents) of production.

The actual effect on total costs is an empirical matter but a sensible hypothesis is that peak spreading does reduce costs with probably more effect on capital than on running costs.

The question then arises of how to obtain a statistical measure of capacity utilization. Several could be suggested but one obvious candidate is Load Factor. Load Factor is usually defined as:

$$\frac{\text{total demand (KW hrs)}}{\text{peak demand (KW x 8760 hours)}}$$

where the peak demand is measured over one half hour on the day of maximum system demand. This measure has 2 drawbacks - firstly it is not available for most states over a period of years and, secondly, it is liable to large random weather disturbances. A less usual but more helpful measure is what might be called Capacity Load Factor (CLF):

$$\frac{\text{total demand (KW hrs)}}{\text{Capacity (KW x 8760 hours)}}$$

Using this measure our basic cost relationship is:

$$uc(i) = a_o + a_1 X^*_t + a_2 V^{(i)}_t + a_3 \; CLF_t + u_{it} \quad \ldots\ldots\ldots\ldots (3)$$

The first term represents scale effects according as $a_1 \gtreqless o$, the second term represents exogenous factor price influences with $a_2 > o$ as the a priori assumption and the third term represents peak shifting influences, $a_3 < o$.

PEAK SPREADING

In developing this notion of measuring planned costs as the predictions from the fitted form of equations, we have been compelled to take account of the effect of peak load variations on unit factor costs. In principle we hypothesise $a_3 < o$ but its actual value is of interest because the estimated coefficient for the i^{th} factor

$$\hat{a}_3^{(i)}$$

is the effect on unit costs associated with the varying load factors:

$$\text{i.e.} \quad \hat{a}_3^{\,i} = \frac{\Delta UIC}{\Delta CLF}$$

the dollar value of the cost saving from a small degree of peak spreading. Hence in the fitted cost regressions the magnitudes of the $\hat{a}_3$ coefficients are of prime importance: they measure the opportunity costs of not attempting to spread the peak and may give us a clue to the importance of the peak loading problem.

VARYING THE SPECIFICATION TO TAKE ACCOUNT OF SERIALLY CORRELATED RESIDUALS

It was noticeable that with some of the equations, (especially those for interest and depreciation costs) problems of serial correlation in the residuals began to arise, as indicated by low values for the

Durbin-Watson statistic. There are two main ways of tackling this problem: firstly, by re-estimating the equation by some type of autoregressive least squares until serial correlation disappears or secondly, by re-specifying the equations. The latter approach tackles what may be the real underlying cause of the problem: the omission of relevant explanatory variables.

After several experiments, it became clear that the simplest way of removing the autocorrelation was to include lagged values of the dependent variable (a procedure fraught with danger as explained below). It seems to be the case that the lagged dependent was the most suitable proxy for omitted variables and also for those variables only poorly represented by the data measuring them. (It is on this basis as a proxy variable rather than on any rationalization of "partial adjustment processes" that the lagged dependent was included.)

This does raise severe statistical problems, so we have to be clear what we are doing. We wish to remove autocorrelation by using the lagged dependent; however if autocorrelation <u>remains</u> after the lagged dependent variable is included then the ordinary least squares estimators may be seriously at fault. To be precise: the lagged dependent and any residual autocorrelation together may mean that the explanatory variables and the <u>contemporaneous</u> disturbance term in the euqation are correlated. In this case ordinary least squares does not even give <u>consistent</u> estimators. However if we can be sure that auto-correlation disappears after the inclusion of the lagged dependent variable, then we can assume consistent estimators with ordinary least squares. This is the point at which the second serious problem arises. The usual tests for autocorrelation are not valid when the explanatory variables contain the lagged dependent. Hence our approach has to be as follows:

(1) introduce the lagged dependent and test for autocorrelation of the residuals not by the usual Durbin-Watson test but by Durbin's recent modification designed specifically for the case where the explanatory variables include the lagged dependent (Durbin, 1970).

(2) If the hypothesis of residual autocorrelation is then rejected, then the ordinary least squares estimators are consistent.

For a precise definition of consistency see Johnston, 1972 or Goldberger, 1964. Briefly, it implies that as the sample size increases biases in the regression estimates diminish (though not necessarily to zero) and the variances of the estimated parameters collapse to zero.

When this procedure was adopted, all evidence of residual autocorrelation did disappear and we can therefore proceed with confidence on the basis of the revised estimates. Durbin's test statistic is h

$$h = \left(1 - \frac{DW}{2}\right) \qquad \frac{T}{1-T\hat{v}(b_1)}$$

where DW is the value of the ordinary Durbin-Watson statistic, T is sample size and $\hat{v}(b_1)$ is the estimated variance of the regression coefficient on the lagged dependent variable. On the hypothesis of zero autocorrelation h is standard normally distributed. We therefore reject the hypothesis of residual autocorrelation at the 5% level if $h \leqslant 1.645$.

Summing up therefore in some equations the lagged dependent variable appears but the theoretical results and statistical estimates remain valid.

DATA

The published data on costs is mainly that supplied by annual reports of the state electricity authorities and the Electricity Supply Association of Australia. Factor payments are reported in the four categories:

fuel for generation	(F)
other working expenses	(O)
interest payments	(I)
depreciation charges	(D)

These correspond to the four categories we have already referred to:

unit fuel costs (UFC)

unit other costs (UOC)

unit interest costs (UIC)

unit depreciation costs (UDC)

(the "planned" versions of the costs are of course UFCP, UOCP, UICP, UDCP the predictions from these regressions.)

Let us now take each cost category in turn, examining the influence important in fitting an equation like (3) above. A good guide to the institutional background is McColl (1972) and his views have been quite helpful in formulating and testing the equations.

FUEL COSTS

Fuel costs are the most obvious output related costs, but the relation may go in two ways. Day to day operation of plant would clearly show (on the merit order basis) rising short run fuel cost curves. However on an annual basis, which is all that the published data allow us to consider, scale economies may appear for several reasons. For instance, expanding markets allow Australian authorities to take advantage of technological advances embodied in large overseas generating sets when these are imported. This, in our sample, might be expected chiefly in Victoria, the largest state in terms of consumption. However, Victoria uses a relatively rare technology - brown coal is the generating fuel, since the enormous deposits of the La Trobe valley outweigh the disadvantages of its relatively low energy of combustion (typically about 10 million joules per kilogram compared with 33 million for black coal). Although we might expect little evidence of scale economies in smaller markets, Johnston's early study of British data suggests that scale economies in generation began at fairly low levels of output, (Johnston, 1961).

There is, in addition, one directly observable measure of technical progress in generation, relative thermal efficiency rates, (TE_t). The rates are simply the ratio of energy output to total energy input (using different conversion factors for each type of input).

The chief fuel used in these three States (Tasmania was not examined in this fuel category because of its reliance on hydro-electric power) is coal and coal price is the input cost variable examined. A series on the marginal cost of brown coal was constructed for Victoria on the basis of rather crude figures for residual sales of brown coal by the SECV. South Australia used its own sources of sub-bituminous coal at Leigh Creek and a very useful price series was constructed from the reported delivered value of Leigh Creek Coal in the Electricity Trust of South Australia's annual reports. West Australia relied partly on imports of New South Wales black coal for which a published price series is obtainable.

Several equations corresponding to 3 were tried and the preferred results are shown in Table IV,1. The overall degree of explanation and the precision of the regression coefficients are both high and serial correlation does not seem to have been a serious problem.

Regression Coefficients and Standard errors of explanatory variables

State	X^*_t	Coal Price*	TE_t	CLF_t	$\bar{R}^2$	D.W.
Victoria	-0.0032		-0.0485	-0.0009	0.9633	2.38
	(0.0004)		(0.0519)	(0.002)		
South Australia	0.0015	0.0621	-0.2180		0.9726	1.11
	(0.0022)	(0.0238)	(0.0595)			
West Australia	-0.0201	0.0048	-0.2840	0.0007	0.9803	2.45
	(0.0039)	(0.0019)	(0.1041)	(0.0015)		

*Coal Price: Victoria: index of brown coal prices
South Australia: index of Leigh Creek coal prices
West Australia: index of imported black coal prices (NSW)

Unit fuel costs have fallen in each of these three states over the sample period, as output has grown and as thermal efficiency has shown a positive if erratic degree of improvement. Movements in coal price have moderated the falling costs, and capacity utilization has had no effect in this element of variable costs.

The evidence of scale economies is interesting. In South Australia, output appears not to have significantly affected fuel costs, but in both Victoria and West Australia output enters with a negative sign and is highly significant. In Victoria this slightly drives out the influence of relative improvements in thermal efficiency but this variable has a significantly negative effect as might be expected in both South Australia and West Australia. The input price series for Victoria has not worked at all well and we may have serious doubts about this data. However, the series on coal price for the other States do appear to have shown that rising coal prices have significantly moderated the falling fuel costs.

OTHER WORKING COSTS

There is less reason to believe that other working costs which include administrative overheads and distribution costs will be output related; this will only be the case to the extent that manning costs are important in this category. However, many authorities have blamed rises in other working costs as reflecting the inflationary trends of the rest of the economy and being the source of tariff revisions. Besides planned output, a measure of general price trends, (The Consumer Price Index in that state, CPI_t) and Capacity Load Factor were incorporated into the equations.

Neither Victoria nor Tasmania, in the preferred equations shown in Table IV,2, indicate any relationship between output and other costs. South Australia shows some diseconomies of scale. However the South Australian equations were all extremely suspect and even the preferred equation did not pass a total equation test. The hypothesis that the whole vector of regression coefficients was not significantly different from zero was not rejected at the 95% level on an F test. This leaves only West Australia, from which we may conclude that there were some scale economies to be obtained, presumably from the distribution network.

Leaving aside the suspect South Australian regressions, it is clear that other working costs do significantly reflect inflationary trends in the rest of the economy as represented by CPI_t. Clearly then we should expect movements in these administrative and labour costs to account for some of the changes in tariffs that have appeared over the sample period. However, an offsetting factor is clearly shown by the ability of more even load spreading to reduce costs in this category. Capacity Load Factor has produced negative coefficients in all cases. There is therefore some initial evidence that use of peak off-peak differences in tariffs may have a definite effect in reducing unit costs and therefore reducing tariffs.

Other Working Costs

Preferred Equations

Table IV, 2

Regression coefficients and standard errors of explanatory variables

State	X^*_t	CPI_t	CLF_t	$\bar{R}^2$	D.W.
Victoria		0.499	-0.0068	0.6852	1.49
		(0.093)	(0.0040)		
South Australia	0.013	-0.921	-0.0094	0.2733	1.01
	(0.006)	(0.520)	(0.0040)		
West Australia	-0.024	1.122	-0.0026	0.7479	2.98
	(0.005)	(0.218)	(0.0012)		
Tasmania		0.075	-0.0019	0.7438	1.43
		(0.060)	(0.0009)		

INTEREST COSTS

Both interest and depreciation costs make up the rental price of capital. However to accord with published data we have separated the two influences. If the simple rental price of capital is:

$$P_k = PCE\ (IR + RHO)$$

where PCE is the price of new capital equipment, IR is the interest rate used in the authority's investment decisions and RHO is the rate of depreciation of the capital stock, then we separate the costs as:

$$P_k = PCE \cdot IR + PCE \cdot RHO$$

and the first term is interest costs and the second is depreciation costs.

These are fixed costs, and as we saw above, likely to be much more closely related to the degree of capacity utilization rather than the level of output, although McColl has suggested that capital costs per KW have been falling as larger generating costs have come into use. Several versions of an equation like 3. were tried and no relationship was found between unit interest costs and output. The preferred equations are shown in Table IV,3.

Again South Australia has shown a very poor performance and serious doubts appear to be raised about the cost data reporting in this case. However, the other states all show interesting results with

Interest Costs | Preferred Equations | Table IV, 3

Regression coefficients and standard errors of explanatory variables

State	PCE_t	IR_t	CLF_t	Lagged Dependent	$\bar{R}^2$	Durbin's h
Victoria	0.1197	0.1019	-0.0036	0.5865	0.9547	-0.1821
	(0.0778)	(0.1039)	(0.0019)	(0.1598)		
South Australia (not used)	0.2000	0.2094	-0.0011		0.2834	
	(0.0699)	(0.1591)	(0.0029)			
West Australia	0.0465	0.0009	-0.0027	0.6087	0.8412	-1.1702
	(0.0735)	(0.1167)	(0.0015)	(0.1110)		
Tasmania	0.1029	0.0816	-0.0023	0.4741	0.9032	-1.4765
	(0.0635)	(0.1084)	(0.0011)	(0.1655)		

some quite precise estimates. Changes in the rate of interest series do not appear to have been important (the series used was the short term Commonwealth bond yield). This suggests that changes in the discount rate applied in project analysis may have little influence on final tariffs (though a considerable influence on project timing may remain). McColl suggests that different suppliers of funds to State electricity authorities charge different interest rates. He points particularly to the contrast between borrowing from State Treasuries and borrowing on the open market. A variable reflecting McColl's suggested measure of the degree of reliance on external borrowing was constructed:

$$REBM_t = \left(\frac{\%\ \text{increase in External Capital Liabilities}}{\%\ \text{increase in Original Capital costs of works}} \right)_t$$

but its performance was negligible and hence there is no evidence that different borrowing sources have important relative effects in unit interest costs.

The capacity utilization variable has had a significant influence on this element of fixed costs: CLF_t has the expected negative sign on its regression coefficient and the standard error is very low in most of the cases. Here is clear evidence that movements in load factor have had significant effects on costs and hence we may strongly argue for peak off-peak tariff differences on this basis.

DEPRECIATION COSTS

As shown above we expect depreciation charges to vary with the price of capital equipment, PCE_t, and with the rate of decay assumed for the capital stock, RHO_t. This latter parameter is usually derived on a rule of thumb basis according to taxation provisions, lifetime of assets assumptions and other factors; in the strict operation of benefit-less-cost minimization we saw how it would vary with the degree of capacity utilization. It is not however directly observable and an instrumental variable has to be used instead. Since we know that State Electricity Authorities have used rule of thumb depreciation methods, an instrumental variable based on the rate of growth of stocks may easily be derived, because of the strict relationship between the rate of decay of stocks and the rate of growth of stocks, when asset life is assumed constant.

Suppose the rule of thumb adopted is one of declining balance so that the depreciation amount is simply the depreciation rate (RHO) times the capital stock; re-writing RHO as r for convenience, declining balance depreciation writes off

$$rK(t) \qquad (4)$$

units of capacity each period. Now the gross investment identity is:

$$I(t) = \dot{K}(t) + rK(t) \qquad (5)$$

i.e. gross investment is net capital accumulation plus depreciation investment. Now suppose capacity is growing at an average rate of i per period

$$K(t) = K(o)e^{it} \tag{6}$$

Consider the initial investment in capacity: $I(o)$; if a unit of capacity lasts for n years then depreciation in period n will replace this initial investment

$$rK(n) = I(o)$$

or in general

$$rK\,(t+n) = I(t) \tag{7}$$

Now combining (5),(6) and (7) we have

$$rK(o)e^{i(t+n)} = iK(o)e^{it} + rK(o)e^{it} \tag{8}$$

Hence we may write:

$$r = \frac{iK(o)e^{it} + rK(o)e^{it}}{K(o)e^{i(t+n)}}$$

giving:

$$r = \frac{i}{e^{in} - 1} \tag{9}$$

and from (9) we may easily show:

$$\frac{dr}{di} < 0$$

under general conditions. This simply says that the faster the rate of growth of stocks, of a given life, the slower the rate at which the volume of stocks decays. Similar results may be obtained using other rules of thumb. Thus we have the general relationship

$$r = f(i,n)$$

with $f_i < 0$, where n is the lifetime of capital equipment.

On this basis the instrumental variable for RHO is taken as

$$\frac{\Delta(\text{Generating Capacity})_t}{(\text{Generating Capacity})_{t-1}}$$

and the expected sign on the regression coefficient is negative. PCE_t and CLF_t are the other main explanatory variables concerned.

However, we must be careful to note that different rules of thumb, assumed asset life and coverage of assets to be depreciated may have been adopted over the sample period. Based on McColl's discussion, the following timetable was drawn up:

	1. Change in assumed asset life	2. Change in depreciation rule applied
Victoria	1965/66	1957/58; 1960/61
South Australia	-	1962/63
West Australia	?	?
Tasmania	-	-

Dummy variables were constructed to represent the changes in this timetable and included in the depreciation costs regressions reported in Table IV,4. DUM 1_t represents changes in asset life and DUM 2_t

represents changes in the depreciation rule applied. On the whole, the equations have fitted reasonably well.

The instrumental variable representing the depreciation rate, RHO_t, performs well in all cases.

An interesting feature of the results here is that changes in the depreciation rule have been important enough to obscure the relationship between this element of fixed costs and capacity utilization. Where $DUM\ 1_t$ or $DUM\ 2_t$ appear in an equation CLF_t becomes unimportant; only in the cases of Tasmania and West Australia where the same depreciation conventions have been applied over the whole sample period, is there clear evidence of the effect of spreading load in reducing fixed costs. CLF_t is significantly negative in these cases. The important dummy variable is $DUM\ 2_t$ in the other cases; it is particularly significant in the case of Victoria where there have been changes not only in the depreciation convention (the adoption of straight line rather than sinking fund conventions) but also in the extent to which the distribution network has been depreciated in these reported changes.

Although no depreciation convention changes were reported for West Australia and Tasmania, two factors were examined to determine whether or not there may in fact have been changes adopted significant enough to affect costs. One factor, the performance of CLF_t has already indicated that no obscuring changes have been apparent; secondly, the residuals in the regression, showed no outstandingly large positive or negative values in any year of the sample period.

Depreciation Costs — Preferred Equations — Table IV, 4

State	PCE_t	RHO_t	DUM 1_t	DUM 2_t	CLF_t	Lagged dependent	$\bar{R}^2$	test for auto-correlation (DW or Durbin's h)
Victoria	0.2571	-0.0368	0.0991	0.6156			0.9914	2.15 (DW)
	(0.0397)	(0.0319)	(0.0913)	(0.0632)				
South Australia	0.1834	-0.0763		0.1657			0.9046	2.03 (DW)
	(0.0305)	(0.0510)		(0.1176)				
West Australia	0.0335	-0.1204			-0.0028	0.6132	0.9245	-0.3004 (h)
	(0.0364)	(0.0355)			(0.0010)	(0.0762)		
Tasmania	0.0058	-0.0322			-0.0013	0.1547	0.6075	+0.8543 (h)
	(0.0069)	(0.0165)			(0.0003)	(0.1478)		

It seems apparent therefore that, in the absence of changes in accounting conventions, a clearly significant relationship exists between fixed costs on the one hand, capacity utilization on the other.

This latter fact, the importance of spreading load in reducing overall costs and hence tariffs must be regarded as a significant argument for the use of peak versus off-peak pricing differentials, advertising of off-peak facilities, (like night-storage heaters) and other marketing strategies to spread load more evenly.

DOLLAR ORDERS OF MAGNITUDE ON THE PEAK LOADING PROBLEM

One outstanding result of the regressions is the strongly significant effect of capacity load factor variations in affecting unit costs. The regression estimates permit us to put orders of magnitude on these cost savings. Tabel IV, 5, presents the main results and the total effect summed over different factor categories.

It seems that the cost savings are not trivial. To get some idea of the impact of these magnitudes we can pick, for example, Victoria or Tasmania. Capacity cost savings alone total \$ 442,000 and \$ 177,000 respectively from a change in load factor of one percentage point (by this I mean a change from, for example, 48% to 49% - this is what I call a one percent change). The Tasmanian case has some interesting applications.

At the time of writing the "Lake Pedder" controversy has had some impact on electricity policy discussion. The Tasmanian Hydro

Table IV, 5 The Cost Savings from Spreading Load

(a) Regression Coefficients $\Delta UIC/\Delta CLF$

	fuel	other	interest	depreciation
Victoria	-0.0009	-0.0068	-0.0036	-
South Australia	-	-0.0094	-0.0011	-
West Australia	-	-0.0026	-0.0027	-0.0028
Tasmania	-	-0.0019	-0.0023	-0.0013

(b) Dollar values of the cost savings from raising CLF by one percentage point: ($ '000)

	fuel	other	interest	depreciation
Victoria	110	834	442	-
South Australia	-	383	45	-
West Australia	-	56	58	60
Tasmania	-	93	113	64

Electric Commission - arguing on the basis of a steadily rising maximum demand - proposed (and are carrying out) a programme of hydro-electric scheme construction in and around Lake Pedder a nature reserve with conservation value. Without making value judgements the economist can point out:

(a) the growth of maximum demand may not be immutable; it may be variable within the control of H.E.C. pricing policy.

(b) the capacity cost savings above suggest that raising CLF for Tasmania by 5.2% could save an annual capacity cost sum of one million dollars. The annual equivalent charge of the Lake Pedder scheme has been variously estimated by the newspapers, critics and public bodies as between 0.9 and 3.6 million dollars (according to assumed life and total cost at an eight per cent discount rate).

Hence the annual costs of the Lake Pedder scheme (on the lower estimate) can be entirely saved by a 5% shift in load factor (a 5% degree of peak shifting).

It looks on this basis as if the potential savings due to CLF variations are within the range at which they can have significant impact in Social Welfare terms.

However, although we have provided some sort of measure of the dollar value of solving the peak load problem, this still leaves unanswered the question of whether tariff policy itself can affect the peak - in particular can raise CLF.

It is this final question that the remaining sections of this part of the thesis are addressed.

CHAPTER V

TARIFFS AND LOAD FACTOR

The point has been made that charging capacity costs relatively more to peak demand not only is a more efficient method of allocating costs (in the economic sense of the word efficiency) but may also be effective enough to persuade consumers to raise the overall load factor of the generating system. Although electricity authorities have been reluctant to adopt peak load pricing policies, it may nevertheless be true that within the limits of traditional tariff structures some manipulation has occurred. This is difficult to test because of the complexity of Australian tariff structures. However, the traditional tariffs do have certain uniform categories within which tariff rates have been varied. With a large enough sample these tariffs provide a useful set of qualitative variables which may be effective in spreading out the peak of electricity demand. In simple terms, the hypothesis is that the attributes or characteristics of a given tariff structure may affect the "degree of peak" in the demand that results.

To formalize a model we can conceive of a measure of degree of peak - in the form of a certain load factor - being the result of consumer reactions to various characteristics of the tariff involved:

$$L^*_t = \sum_{j=1}^{J} a_j C_{tj}$$

where L^*_t is the desired overall load factor depending on the different characteristics of the tariff structure, $C_{t1} \ldots C_{tJ}$, at time t.

There ought to be an allowance for a time lag in the adjustment of actual load factor to this desired overall level, described by the difference equation

$$L_t - L_{t-1} = W\,(L^*_t - L_{t-1})$$

where, by assumption: $0 \leqslant W \leqslant 1$.

Consequently actual load factor - the measure of the peak character of any demand schedule - is given by

$$L_t = W \sum_{j=1}^{J} a_j C_{tj} + (1-W)\, L_{t-1}$$

The initial experiments to relate load factor to tariffs in this study consists of fitting an equation of this type to a pooled sample of time series and cross section data for our four states.

The characteristics of the prevailing tariff structures are qualitative variables conveniently represented by the following type of dummy variables:

(i) $C_{tj} = 1$ if the tariff structure has option j at time t

(ii) $C_{tj} = 2$ if option j is "emphasised" (e.g. by application to an extended number of consumer categories) at time t

(iii) $C_{tj} = 0$ if option j is absent from the tariff structure at time t

A three way classification seemed preferable because while some tariff structures offer, for instance, off-peak prices as one of a number of categories available to a limited number of consumers, others offer off-peak prices as an option to almost every tariff, applicable to almost every consumer.

Of course, it then becomes a matter of judgement deciding exactly what values to give to different tariff characteristics at any one time; however as general as possible a classification was used consistent with keeping the model in a workable shape.

Two statistical comments are in order before describing the experimental results. The dummy variable trap was avoided by constraining the regression plane to pass through the origin, and, secondly, the experimental procedure does not allow tariff characteristics which remained unchanged over the sample period to be included. This is because the result of including a variable that was always 1 or 0 in value would have meant a linearly dependent set of columns in the data matrix causing ordinary least squares regression to break down.

As a consequence of the last point, the lagged dependent term picks up the underlying effects of the non-changing parts of the tariff structure while the variables C_{tj} isolate the influences on the movement of L_t associated with variations in the tariff structure.

In the final analysis four tariff characteristics were isolated as having appeared disappeared or changed in nature over the sample. Three of these were:

the provision of a time of day tariff option : C_{t1}

the provision of a maximum demand option : C_{t2}

the provision of a night storage heating option : C_{t3}

The fourth qualitative variable consisted of a measure of tariff alterations. Each time that part of the tariff structure of the state in question was altered (a rate was raised or amalgamated with another rate etc.) this fourth dummy variable increased in value by 1. It therefore measures the effect on load factor of any tariff alterations whatever the motivation for them.

As mentioned above the sample is a pooled time series cross section mixture so that tariff restructing variable is in fact split into 4 variables one each for Victoria, South Australia, West Australia and Tasmania; e.g. when the Victorian section of the L_t data ends a new dummy variable for the next state starts to increase as tariffs are changed.

Thus the final estimating equation is:

$$L_t = \beta_1 C_{t1} + \beta_2 C_{t2} + \beta_3 C_{t3} + \beta_4 \Delta TS_{1t} + \beta_5 \Delta TS_{2t} + \beta_6 \Delta TS_{3t} + \beta_7 \Delta TS_{4t} + \beta_8 L_{t-1}$$

where C_{tj} is a dummy variable representing a tariff option and ΔTS_{ti} is a dummy variable representing some change in a State's tariff structure.

Various versions of the equation are fitted to the pooled sample of states over the period 1953-1971, and the results appear in Table V,1 below. The lagged dependent - indicating the overall influence of the "permanent" parts of the tariff structure - dominates the explanation but the signs on the qualitative variables are of great interest.

On the basis of best overall fit equation (3) is to be preferred. The results provide the following conclusions:

(i) CLF adjusts slowly to the measure of tariff structure

(ii) the effect of the provision of an off-peak tariff option is to raise the overall load factor. The off-peak option C_{t3} is the cheap night rate for storage heating.

(iii) all the dummies representing tariff structure alterations show a positive relationship with load factor except for the variable representing Tasmania. It would appear that the impact of tariff alterations in Tasmania - whatever their real purpose - has been to lower load factor.

From looking at the other equations (e.g. equations (1) and (7)) the variable C_{t1} - provision of a general time of day option - also raises load factor but by a smaller degree than the particular option of a night storage heating rate; (the size of the regression coefficients in equation (7) can be directly compared since all the

CLF Regressions

Pooled sample 1953-1971, States of Victoria, Tasmania, South Australia and West Australia.

Partial regression coefficients — dependent variable CLF_t

Equation No.	CLF_{t-1}	C_{t1}	C_{t2}	C_{t3}	ΔTS_{VIC}	ΔTS_{SA}	ΔTS_{WA}	ΔTS_{TAS}	$\bar{R}^2$
1.	0.96*	1.30			0.16	0.39*	0.37	-0.19	0.66
2.	0.98*		0.66		0.08	0.20	0.26	-0.10	0.66
3.	0.95*			2.05*	0.07	0.43*	0.18	-0.02	0.66
4.	0.96*	1.28	0.02		0.16	0.39*	0.37	-0.19	0.66
5.	0.95*	0.20		1.84	0.08	0.44*	0.20	-0.05	0.66
6.	0.95*		-0.37	2.38*	0.08	0.45*	0.22	0.05	0.66
7.	0.95*	0.25	-0.39	2.13	0.09	0.46*	0.25	0.02	0.65

* regression coefficient exceeds its standard error.

variables are dummies measured in integer units). Attempts to influence load factor by the use of maximum demand charge tariff options seem to have been unsuccessful - equations (6) and (7) even show a negative sign on C_{t2}.

So far the experiments of this and the previous sections seem to give us two conclusions:

- load factor variations do have a cost saving impact.
- load factor itself appears to react positively to the provision of off peak charges.

It would be helpful therefore to have some measure of the magnitude of load factor variations obtainable from the provision of off-peak rates. For this task, a second set of experiments is necessary.

LOAD FACTOR VARIATIONS AND OFF-PEAK TARIFFS

The aim of this section is to measure the magnitude of load factor variations obtainable when various off-peak tariff rates are varied. In precise terms we wish to calculate an elasticity of load factor with respect to variations in the ratio of off-peak to peak tariff rates.

$$E = \frac{\Delta CLF}{\Delta PI} \cdot \frac{PI}{CLF}$$

where PI is a variable measured by the ratio of a given off peak tariff to its on peak (daytime) alternative. If peak spreading does occur we expect this elasticity to be negative:

$$E_{CLF,\ PI} < 0 \qquad \text{by hypothesis}$$

Thus a rise in the night rate relatively to the day rate is expected to reduce load factor, while a fall in the night rate relatively to the day rate is expected to raise load factor. Naturally the hypothesis is that the chain of causation runs from PI to CLF.

However there is a possibility that there may be a feedback in the direction of CLF to PI. It has already been established that variations in CLF can affect costs. A rise in CLF may have enough effect on the ratio of capital to running costs to persuade an electricity authority to change a previously established relationship between off-peak and peak tariffs. The model is really one of simultaneous determination of CLF and PI allowing for interdependency between these variables. The essence of such a model can be formulated very simply:-

Equation (1)

$$CLF_t = \alpha_1 + \beta_1 PI_t + \delta_1 \left[\text{predetermined variables} \right]_t$$

Equation (2)

$$PI_t = \alpha_2 + \beta_2 \, CLF_t + \delta_2 \left[\text{predetermined variables} \right]_t$$

Our hypotheses about the coefficients would now be as follows:

$$\beta_1 < o$$

measuring the effect of the tariff variations on load factor - the effect whose magnitude is of prime interest.

$\beta_2 \gtrless o$ according as cost changes due to varying load factor are incorporated in tariff structures.

Unless this simultaneity and interdependence is taken account of, consistent estimates of the magnitudes we are interested in cannot be obtained.

There are several ways of estimating the equations of a simple model like this and for the three or four versions used in the experiment, two stage least squares (2SLS) was chosen. For a description see Johnston (1972 chapter 13).

In different versions the "predetermined variables" (which may include past values of the "jointly dependent" variables) used were:

equation (1) CLF_{t-1} : TIME

equation (2) CR_t : the ratio of running to capacity costs per unit of output at time t

The model can only be said to have been successful in terms of the expected behaviour of the signs on the coefficients for one state: Victoria. The results here are of some interest however. It is not surprising that Victoria should be a relative success in this respect since it is (or would claim to be) the state with longest allegiance to the idea of time of day tariffs.

Of the various candidates for use as PI_t two were chosen:

$PI(1)_t$: the ratio of industrial and commerical night rate to the first block of the equivalent daytime running charge.

$PI(2)_t$: the ratio of the night rate for water heating for residential consumers to the equivalent residential daytime running charge.

From the 2SLS estimates of equation (1) the following elasticities were calculated (at the data means):

$$E_{CLF,\ PI(1)} = -4.8$$

$$E_{CLF,\ PI(2)} = -6.1$$

To obtain an appreciation of what these really mean consider the following hypothetical examples:

(a) In 1971 capacity load factor for Victoria was 49.7% and the relation between night and day rates for PI(1) (industrial consumers) was:

night	day
1.00 ¢ per KWh	8.00 ¢ per KWh

Then if the night rate was reduced to 0.90 ¢ per KWh and the day rate stayed the same the model predicts that 1971 load factor would have risen to 50.0%.

(b) Again taking 1971 as an example the night and day rates making up PI(2) were: (residential consumers)

night	day
1.05 ¢ per KWh	2.21 ¢ per KWh

Reducing the night rate to 0.95 ¢ per KWh and leaving the day rate unchanged at 2.21 ¢ per KWh would have raised load factor (according to the model) to 50.1%.

In conjunction with the earlier measures of cost savings due to load factor variations, such rises in load factor are by no means negligible.

SUMMING UP THE WORK ON COSTS AND TARIFFS

Very briefly, the overall impressions resulting from the analysis and experiments just described can be set out.

To begin with we discussed the extensive literature on theoretical principles of setting electricity tariffs. The key to the economist's view is that "timing of demand" is the basic determinant of costs and should be reflected in "time of day" prices if marginal cost pricing is adopted. The essence of what might be called the engineering approach is to distinguish between capacity and running costs and to charge separate prices accordingly. However there is an intervening area in which one scheme becomes very closely related to the other. In addition there are several suggested forms of discrimination among consumer groups.

The first experiments showed that all types of pricing principles make an appearance in Australian electricity tariffs but in experiments to correlate particular tariff categories with particular cost categories, it was found that what the theoretical principles recommended in relating tariffs to running and capacity costs were not practiced in general. Indeed sometimes the relationships were exactly opposed to the theoretically correct ones.

Moving on, we examined whether the "timing of demand problem" had a significant cost dimension. It was found that spreading peak demand into off peak periods - as measured by rises in capacity load factor - did have a measurable and arguably important impact on unit costs.

This lead to posing the final question: given that load factor variations have a cost impact, can such variations in timing of demand be achieved by tariff alterations? A first set of experiments indicated that capacity load factor was positively related to certain "time of day" characteristics of the tariff structure and further experiments suggested - for the State of Victoria - some orders of magnitude for these effects.

It is on the basis of this analysis and observation that it can be argued that pricing policy is of fundamental practical importance in resource allocation in electricity supply in Australia.

APPENDIX TO PART I

COST INFLATION AND PEAK LOAD PRICING

The effects of cost inflation on peak and off peak prices can be examined in the simplest two period model which has appeared several times in the literature - for example Littlechild (1970). The essential result is that for a social optimum, price will exceed marginal running cost in periods when demand is high by an amount which just restricts demand to capacity; summing all these differences between price and marginal running cost we obtain a term equal to marginal capacity cost. We need only examine the case in which the demand cycle is broken into two periods. In symbols our solution is:

$$p_1 = c + v_1 \qquad \text{..........} (1)$$

$$p_2 = c + v_2 \qquad \text{..........} (2)$$

$$v_1 + v_2 = b \qquad \text{..........} (3)$$

$$v_1 \geqslant o \qquad v_2 \geqslant o$$

Where p_1 and p_2 are prices charged in periods 1 and 2 respectively, c is marginal running cost in those periods and b is marginal capacity costs; v_1 and v_2 are the differences between price and marginal running cost in each period and mathematically represent the Kuhn-Tucker multipliers on the existing capacity constraints.

We may begin the analysis with the simplest possible case and assume that the only inflation we observe is a steady rise in marginal running costs. (To keep the notation simple we shall measure all cost changes as being net of technical progress.) This case reflects what might happen if there were a newly negotiated wage structure in the industry.

Taking partial derivatives of equations (1) and (2)

$$\frac{\partial p_1}{\partial c} = 1 = \frac{\partial p_2}{\partial c}$$

i.e. peak and off peak prices rise by the same absolute amount.

In percentage terms:

$$\frac{1}{p_1} \frac{\partial p_1}{\partial c} = \frac{1}{c + v_1}$$

$$\frac{1}{p_2} \frac{\partial p_2}{\partial c} = \frac{1}{c + v_2}$$

Suppose period 1 is the off peak period, so that

$$o < v_1 < v_2$$

then

$$\frac{1}{p_1} \frac{\partial p_1}{\partial c} > \frac{1}{p_2} \frac{\partial p_2}{\partial c}$$

and clearly off peak price ought to rise by a greater percentage than peak price. The reason is simply that we are only allowing inflation to affect running costs and these have a greater weighting in off peak than in peak price.

Obviously then there is some justification for the proposed new tariff structure having a relatively lower peak to off peak ratio.

However, this is only the most oversimplified case. Inflation may be apparent in capacity costs as well as running costs and this may alter the results considerably. To investigate this we can specify our expression for capacity cost more formally. Capacity cost will depend on several factors - the price of new equipment, the rate of interest, the economic life of equipment. A simple case is that in which the capacity is capable of producing the same output year after year until it stops operating completely. We can then consider associating a constant amount of capacity cost with each year of operation and therefore obtain an expression for capacity cost as the price of new equipment (K) multiplied by an appropriate annuity rate

$$b \quad = \quad \frac{i(1+i)^n}{(1+i)^n - 1} \quad K \qquad \ldots\ldots\ldots\ldots (4)$$

Where i is the rate of interest (the cost of capital and the discount rate used by the electricity authorities) and n is the life of the equipment. The period we are interested in is very short relatively to the life of the equipment so that we can simplify the expression in equation (4) by considering

$$b \quad = \quad \lim_{n \to \infty} \quad \frac{i(1+i)^n}{(1+i)^n - 1} \quad K \quad = \quad iK \qquad \ldots\ldots\ldots\ldots (5)$$

Using equation (5) our simple peak and off peak pricing model becomes:

$$p_1 = c + a_1 \text{ (iK)} \qquad \text{..........} \quad (6)$$

$$p_2 = c + a_2 \text{ (iK)} \qquad \text{..........} \quad (7)$$

$$a_1 + a_2 = 1$$

$$a_1 > 0, \quad a_2 > 0$$

and we have replaced v_1 and v_2 by terms expressing the proportion of marginal capacity cost charged to each period; i.e. a_1 (iK) in period 1 and a_2 (iK) in period 2.

We may now examine the case of cost inflation in both running costs and in the price of new equipment. Take total differentials of equations (6) and (7) and express in percentage terms:

$$\frac{dp_j}{p_j} = \frac{c(dc/c) + K\, a_j\, i\, (dK/K)}{c + a_j\, (iK)} \qquad \text{..........} \quad (8)$$

for j = 1,2.

The effect on price for each period will reflect the weighted average of inflation rates in running and capacity costs. The most straightforward case is

$$dc/c = dK/K = g$$

i.e. the rate of inflation is the same for running costs and capacity costs; substituting in equation (8)

$$\frac{dp_1}{p_1} = \frac{dp_2}{p_2} = g$$

peak and off-peak prices rise by the same percentage - the common rate of cost inflation - and there is no basis for altering the peak off-peak price differential.

Summing up so far, we can see that our first case of inflation only in running costs is a special category of the general rule embodied in equation (8):

(i) the effect on price is to raise it by a percentage which is a weighted average of the percentage rates of inflation in running and capital costs.

(ii) if the rate of inflation is the same for both categories of cost the peak off-peak price ratio does not change.

(iii) off-peak price rises by a greater percentage than peak price only if inflation in running costs outpaces inflation in capacity costs.

Hence to provide a justification for tariff changes we have to assume that case (iii) applies and that its effect is sufficiently severe for it to outweigh the social costs imposed by a reduction in the ratio of peak to off-peak price.

A numerical illustration of the results implied by equation (9) is shown in table 1.

Table 1.

Illustration of the results of equation (9)

assumptions: $dc/c = 0.10$ $c = 2$ $i = 0.10$

$dK/K = 0.05$ $K = 100$

a_1	a_2	dp_1/p_1	dp_2/p_2
0.0	1.0	0.10	0.058
0.3	0.7	0.07	0.061
0.5	0.5	0.065	0.065

Table 1 illustrates a case where the rate of cost inflation in running costs is double the rate for capacity costs. As we move from a case of very high demand in period 2 and very low demand in period 1 (first row of table) to a case where demand is the same in both periods, the amount by which the percentage rise in off peak price exceeds the percentage rise in peak price diminishes rapidly.

However, it can be argued that another case remains. We have not allowed inflation to appear in the rate of interest term. In normal discounting analysis an approximate method of taking account of the decrease in value of future sums due to inflation rather than due to time is simply to add the expected rate of inflation to the discount rate

$$i = r + q$$

where r is the real rate of interest and q is the percentage rate of inflation expected over the time period in consideration. If it is the case that private industry is consistently adding an inflation premium q to its own discount rates, then sooner or later the public utilities /government may wish to change their discount rates to remove the bias in public sector resource allocation. We ought therefore to discover the effect on peak prices and off-peak prices of simultaneous inflation in capital and running costs <u>and</u> an alteration in the discount rate applied.

If we allow the discount rate to rise by a small factor from i to $(1 + \Delta)i$ we obtain from equation (8) the expression:

$$\frac{dp_j}{p_j} = \frac{c(dc/c) + a_j\ (iK)\ (dK/K) + \Delta\ i\ a_j\ K}{c + a_j\ (iK)}$$

If we assume

$dc/c = dK/K = g$

we obtain:

$$\frac{dp_j}{p_j} = g + \frac{a_j\ K\ \Delta\ i}{c + a_j\ (iK)} \qquad \text{......... (9)}$$

To investigate the effects of this expression consider the numerical example shown in table 2.

Table 2.

Illustration of effects of cost inflation and an altered discount rate.

assumptions: $g = 0.15$ $i = 0.10$ $c = 2$
$i = 0.05$ $K = 100$

a_j	$dp_j/p_j = m + (a_j\ K\ \Delta\ i/c + a_j\ i\ K)$
0.0	0.150
0.2	0.400
0.4	0.483
0.5	0.525
0.8	0.550
1.0	0.567

The table illustrates that the effect of raising the discount rate from ten percent to fifteen percent is a very large rise in the capacity cost element in prices: the higher is a_j (indicating a high peak demand) the higher is dp_j/p_j.

The nature of the effects can be seen in figure 1; quadrants I and III plot the expression for dp_j/p_j against a_j; quadrant II indicates the variation in a_j - e.g. for $a_1 = 1$ and $a_2 = 0$, all demand is concentrated in period 1; for $a_1 = 0.5 = a_2$ demand is spread equally over both periods. Quadrant IV plots the relative effects on dp_1/p_1 and dp_2/p_2 for different values of a_1 and a_2. The clear result is that when

FIGURE 1

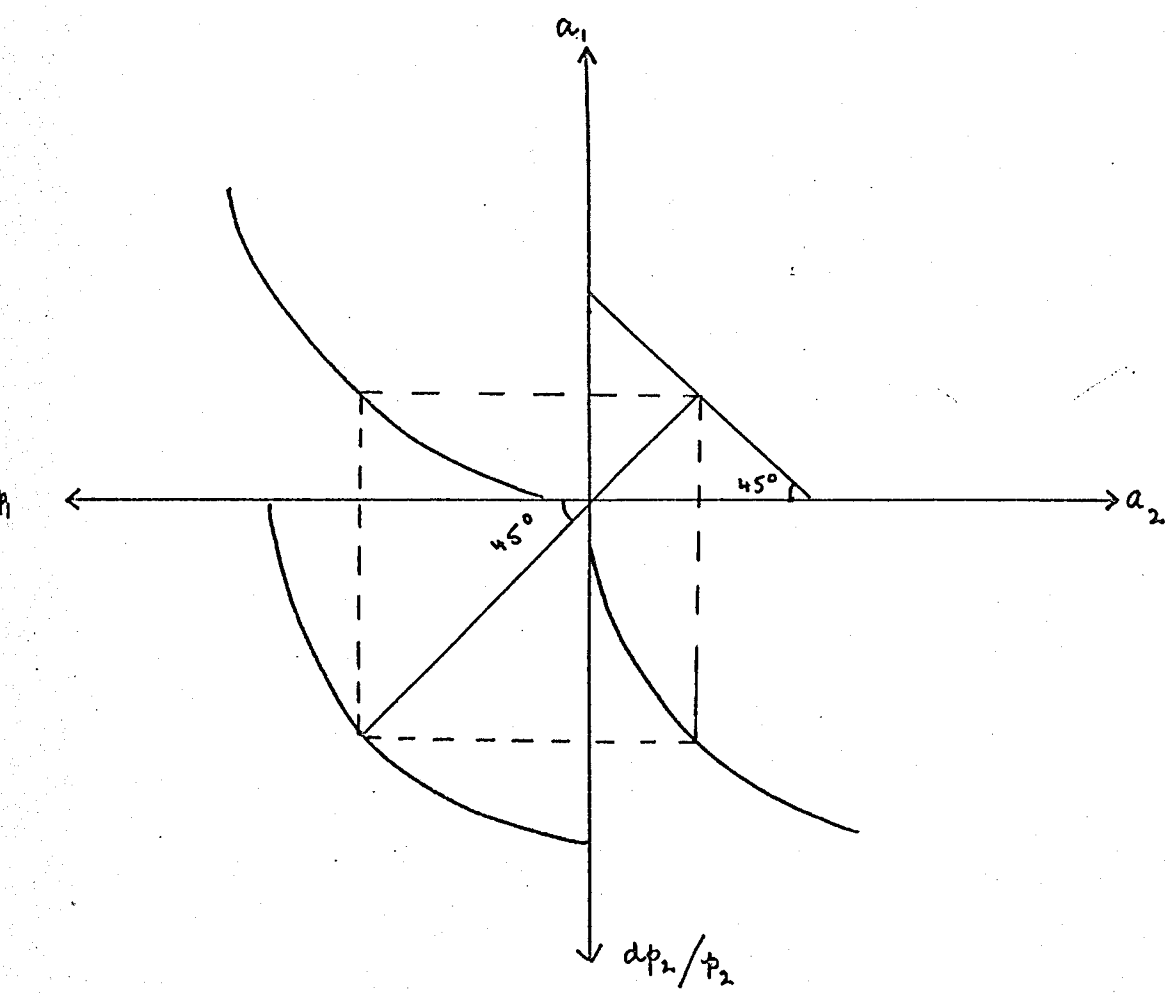
The effects of cost inflation and an altered test discount rate on peak and off-peak prices:

dp_1/p_1 = percentage change in period 1 price.

dp_2/p_2 = percentage change in period 2 price.

a_1 = proportion of capacity costs allocated to period 1.

a_2 = proportion of capacity costs allocated to period 2.



a rise in the discount rate is allowed for, the effect of inflation is to raise peak prices relatively to off-peak prices. For example, if period 1 is off-peak demand and twenty percent of capital costs are allocated to this period then from table 2, (row 2) we observe

$$\frac{dp_1}{p_1} = 0.40$$

and since the other eighty percent of capital costs are allocated to period 2 we note from row 5 of the table:

$$\frac{dp_2}{p_2} = 0.55$$

Off-peak price rises by forty percent while peak price rises by fifty-five percent.

We ought to be clear why this has happened. Raising the discount rate raises the capital charges (capacity costs) associated with each period because these capital charges reflect the reduction for each period in the net present value of the future earnings of the equipment in question. Any rise in the opportunity cost of new investment (i) makes this reduction greater and hence adds to each period's capital charge. Since capital charges have an additional increase, peak prices which bear the brunt of capital charges will rise proportionately more than off-peak prices. The nature of this view of capital charges is well discussed by Salter (1966) (page 19).

It can be seen therefore that we have a variety of results depending on the assumptions used:

(a) if running costs show greater inflation than capacity costs we would normally expect off-peak prices to rise by a greater percentage than peak prices.

(b) where running costs and capacity costs rise at the same rate the ratio of peak to off-peak price should not alter.

(c) eventually cost inflation may lead the electricity authorities to evaluate capacity expansion programmes at a new higher test discount rate; this could mean that peak price will rise by a greater percentage than off-peak price.

PART II

INVESTMENT PLANNING MODELS IN ELECTRICITY SUPPLY

CHAPTER VI

A MODEL OF THE INVESTMENT PLANNING PROBLEM

INTRODUCTION Our purpose here is to examine some of the fundamental issues which economists have raised in tackling the investment planning problem in electricity supply. Because theoretical issues are being considered the discussion is at an abstract level; nevertheless all of the models reviewed have, in one way or another, attempted to reflect real-world constraints and all of them have a great deal to say about planning electricity supply in realistic circumstances.

Rather than adopt a chronological investigation, it is preferable to examine different models of electricity supply in the context of one particular approach to the problem. One of the best developed and most explicit models is that given in a series of contributions by Ralph Turvey (Turvey 1968, 1969, 1971a, 1971b). Many of Turvey's results and propositions have received comment (not all favourable) but there seems little disagreement that this series of models has cast a great deal of light on the problems of electricity supply. However, the usefulness of a model's results depend to some extent on the sort of questions that it attempts to answer. Turvey himself has stated that a fundamental result of investment planning models is the indication they give of the optimal price to be charged for electricity. Basically the model provides a clue to the calculation of marginal cost and, in some way or other, marginal cost is expected to enter into the pricing decision. However, a crucial result of the recent analyses is that the old textbook idea of a single easily defined marginal cost is seriously deficient. Putting together the views of Meek (1968) and Turvey (1969) we can begin by noting the following proposition

- investment planning provides a determined optimal path of capacity installation and utilization once the planner's objective function is explicitly stated; as a by-product of the planning exercise information is obtained on the additional cost of small changes in the values of the exogenous variables in the constraints on the objective function; since there are many types of such constraints there will be many types of these additional costs and one or more will be appropriate bases for charging the individuals who are responsible for the exogenous changes -

The two key arguments here are

(a) an argument by Meek that marginal cost pricing is a means and not an end (Meek 1968, p.58); it is a means towards satisfying the planners overall objective;

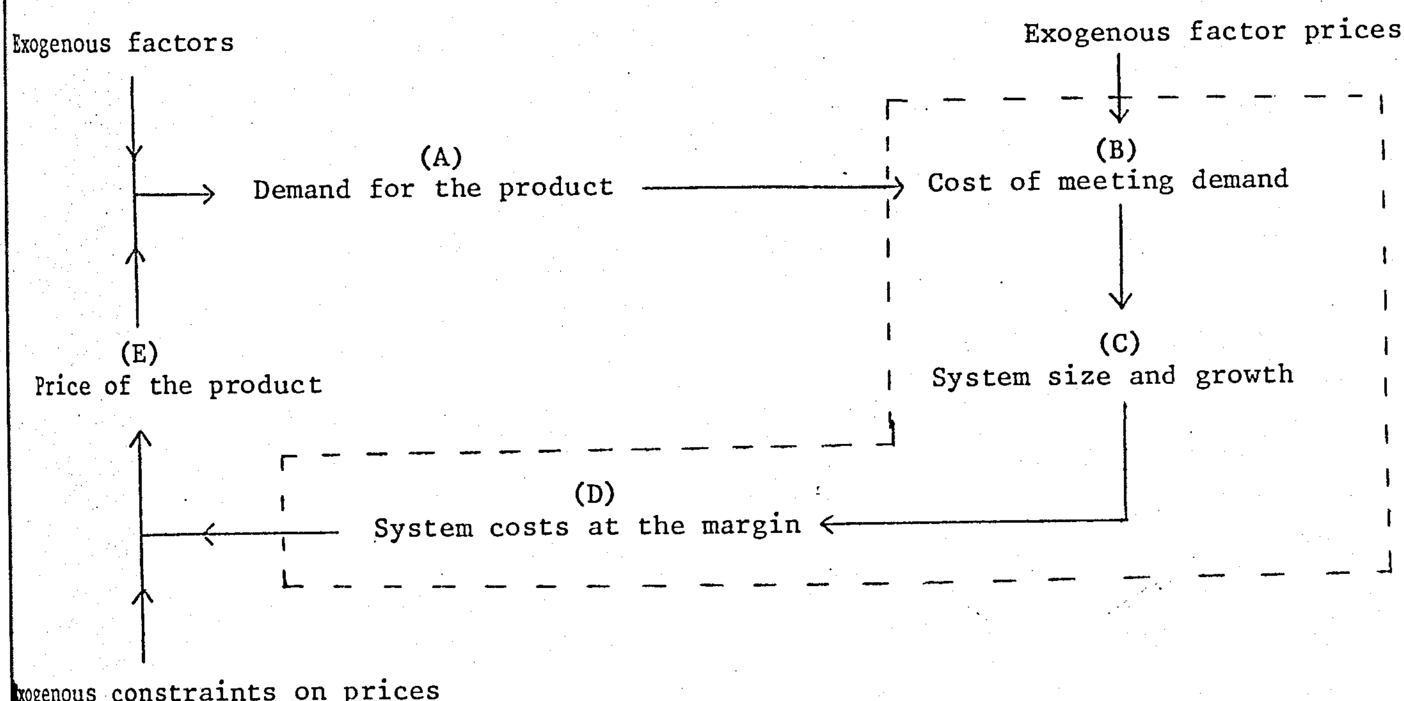
and (b) an argument by Turvey that there are several marginal costs, not just one (Turvey 1969, p.285)

In brief, therefore, we can expect the models to be examined to be concerned with optimising a given objective function subject to constraints and to provide in the solution many different kinds of marginal cost.

SETTING UP THE PROBLEM We are going to examine the investment planning problem in the context at first of a specific series of models so we must initially consider what these models do and what they do not attempt to do. In the introductory chapters of his book on Economic Analysis and Public Enterprises (1971) Turvey sets out very specifically the issues which he - and, as it happens, most of the other writers in this field - has ignored. He is not concerned with income distribution or questions of equity or subsidising certain groups in the community. The problems of externalities are not especially considered and he repeats his belief that in dealing with problems of information "What is not known should be ignored". Although there is a stock of literature on the problem of second best(1) he does not bring this into his discussion of optimal investment planning models, and neither shall we. He believes the planners' objective is and ought to be the maximisation of the difference between social benefits and social costs and social benefits are to be measured by willingness to pay. With this cutting of the Gordian knot of many controversial issues in public enterprise economics he arrives at his statement of the problem of optimising the capacity expansion programme of a theoretical public enterprise which could well be exemplified by any of the Australian electricity authorities.

The planning situation facing the enterprise is modelled in figure VI,1. Demand for the product responds to exogenous factors like the growth of income and to endogenous factors like prices. Having partially determined demand by its pricing policy the enterprise seeks to minimise the costs of meeting that demand, and this optimisation programme not only solves for the optimal path of capacity accumulation but provides the measures of system costs which form the basis of pricing policy along with the exongenous constraints imposed by the planner's political superiors. Turvey makes the empirical judgement that planners know less about the interaction between parts (A) and (E) (i.e. demand and prices) than they do about parts (B) (C) and (D) (system costs and expansion). Rather than tackle the whole planning problem at once he isolates the system cost aspects and optimises on these alone allowing the feedback from prices and demand to have an effect on the second iteration of the cost planning exercise. In terms of computational effort alone this approach is essential, and is explicit or implicit in almost all the writings on the subject with the exception of the peak load pricing literature where

Figure VI, 1 The Planning Situation



The aspects of the problem considered by Turvey are isolated within the box superimposed on (B) (C) and (D).

sufficiently simplifying assumptions allow the whole problem to be tackled at once.

Before we go on to the details of Turvey's models we may note two comments he makes on this approach. Length of lead times and gestation periods in electricity supply are so large that there may be many changes in the exogenous factors between the steps of obtaining an optimal programme and carrying it out. This reduces the likelihood of a convergence towards an optimum in practice; this non-convergence of optimal plans to subsequent optimal practice clearly imposes costs on society. Turvey dismisses any attempt to lower this cost by using a simplified interdependent model: his comment(2) is "life is like that and one might as well try to adjust to it".

Summing up his statement of the problem therefore we can say:

> the planner accepts a set of demand forecasts and tries to minimise the cost of meeting them.

The forecasts could be in probability terms or as certain point forecasts. All of the main models have used the latter formulation and we shall continue to do so.

TURVEY'S SERIES OF MODELS

Turvey specifically tackled the optimal investment problem in electricity supply in his book Optimal Pricing and Investment in Electricity Supply in 1968 but he concentrated there on some of the marginal conditions for local optima. Having determined an overall investment plan he discussed the problems of making marginal adjustments to it. It is only in his later writings (especially 1971a and 1971b) that he sets out his global optimising model.

There are several noteworthy characteristics of an electricity capacity planning model and these can be examined in the context of the Turvey series. The most obvious is that electricity is (except in some unimportant senses) not storable and at the same time subject to severe periodic demand fluctuations. The importance of this is in the way in which marginal system costs are translated into a set of optimal prices, but in the situation of planning minimum cost investment the problem of peaks in supply can be avoided by the very simple device of redefining the time periods involved. Thus instead of considering a set of annual demand forecasts we could specify a set of demand forecasts for the different groups of hours within a year which make up the demand cycle. If we compare Turvey's generalised model (1971a) with his specific programming model for electricity supply (1971b) it is apparent that the structures of the two models are identical, only the time subscripts differ. Hence bearing in mind that the time variables (t) may refer to hours rather than years we need only examine the generalised investment planning model[(3)]. Similarly Turvey's model 1971a allows for only one type of plant while his model 1971b allows for an optimal plant mix to be determined. Nevertheless the underlying structure of the model does not alter and the only difference between the two is the use of an additional superscript to denote plant type. The basic results that we want to consider are unaffected by redefining the time variable or adding superscripts, hence the long run investment planning model is unaffected in abstract terms by ignoring the two considerations of peaked demand and multiple plant types. In a short run model of fixed generating capacity where the problem is optimal despatching these considerations cannot be avoided. The appendix presents a model of this situation[(4)] in the spirit of Turvey's long run model.

We now turn to consider some other characteristics of optimal investment models. We have already stated that the objective function is to minimise the costs - subject to some constraints - of meeting a specified set of demand forecasts. The solution paths are in terms of two variables:

$\{Q^v\}$: Installed capacity at each moment in time

$\{O_t^v\}$: rate of output in period t of capacity (Q^v) installed in period v. (The notation used here means that $\{Q^v\}$ is a sequence (Q^1 Q^2 ... Q^n ...) of installations and $\{O_t^v\}$ represents a sequence of outputs.

In each period total output is the sum of the outputs from each type of capacity already installed.

$$\text{total output} = \sum_{v=o}^{v=t} O_t^v$$

in period t.

The superscript v refers to capacity installed in year v (it could equally refer to capacity of type v).

Associated with capacity and the rate of output of capacity are two costs called capacity and running costs:

c^v : the cost of installing one unit of capacity Q^v built in year v.

r_t^v : the cost of producing one unit of output O_t^v in year t from capacity built in year v.

The expression for total systems costs (C) is therefore:

$$C = (\text{total systems costs}) = \sum_{v=o}^{v=\infty} \left[c^v Q^v + \sum_{t=v}^{t=\infty} r_t^v O_t^v \right] \quad \text{...... (1)}$$

Turvey chooses an infinite horizon in order to simplify the exposition. There are arguments for and against this. An infinite horizon solves the problem of calculating scrap values at the end of a finite horizon; since these scrap values are to be passed on to the next generation anyway having a finite horizon poses as many problems as it solves. However an infinite horizon is not computable in practical terms and analytically raises problems of what the value of the solution should be as we approach infinity. (This "transversality" problem is mentioned again in the appendix.) Nevertheless to obtain the essential results of the model an infinite horizon is just as useful.

The most crucial aspect of the problem is its essentially dynamic nature. Investment programmes are planned over time and a static analysis becomes irrelevant. This has prompted Turvey to argue that it is a <u>historical dynamic</u> concept of marginal cost which emerges from the investment programme making the old textbook static concepts redundant. This is completely true and very important but is not an issue which has yet been resolved as we shall see below. Given that the analysis brings in time in a positive manner the question arises of discounting the future. Turvey's models (and the practice of Australian electricity authorities) accept without demur the necessity of discounting the future and that really resolves the question for us. Nevertheless we should reorganise at least in passing that some commentators argue that in infinite horizon planning models the use of a discount rate other than zero is a defect from a welfare point of view(5).

Having obtained our concept of system costs (C) what can be said about it? One of the most important things is that it allows for what Turvey calls "system interdependence" (Turvey 1969, page 292) so that when an item of new capacity is required its appearance affects the functioning of the rest of capacity. Any cost conclusions relating to one piece of capacity depend crucially on the cost implications of having or not having other pieces of capacity. Turvey's first version of his explicit system planning model (1969) ignored system interdependence. This allowed him to consider in isolation the problem of advancing or postponing the replacement of one item of capacity irrespective of the existence of other capacity. However the rules he derived do have counterparts in the more general model of system interdependence.

In solving his dynamic model Turvey is able to use a static non-linear programming technique by sufficiently splitting up his time variable. Each dated solution value of (Q^v) and (O_t^v) will depend on every other solution value and it is easy to see that this interdependence translated into dynamic terms means that every "current" planning decision depends on a whole series of hypothetical "future" decisions.

The exogenous variables affecting the choice of Q^v and O_t^v are c^v and r_t^v (which we ought to think of as being calculated in present value terms), and in addition the specified set of demand forecasts:

$$\{X_t\} = X_1 \ X_2 \ \ldots.. \ X_n \ \ldots.$$

Turvey makes explicit assumptions about how technological progress affects r_t^v and c^v. For given t, r_t^v and c^v fall with v; i.e. technological progress ensures that the running and capacity costs of later vintage capacity are lower than those of earlier capacity. This corresponds to the assumption that technological progress is embodied in the latest vintage of the capacity acquired. Best practice plants and techniques are those most recently adopted. This idea associated mainly with the names of Salter, Solow and Johansen(6) has received some critical comment on empirical grounds, (Jorgenson 1966, and Gregory and James, 1973) but as the discussion in Chapter VII indicates, it is probably an empirically justifiable assumption for the case of electricity supply.

In addition to r_t^v declining with v, Turvey assumes that r_t^v rises with t for given v. Older plant becomes progressively more costly to run in terms of maintenance and repairs. These two assumptions (embodied technological progress and positive ageing costs) together introduce a non-stationary element into the model, as Turvey recognises (1971b, page 375). The assumption of non-stationarity simply means that, apart from the dependence of (Q^v) and (O_t^v) on time, the variable (t) - time - enters explicitly into the objective function since r_t^v and c^v can both be replaced by expressions in t. If this was not the case then at least one of the optimal paths would mean that O_t^v and Q^v would settle down to stationary values unchanging for the rest of time, and the problem would be said to be "autonomous". It is therefore the crucial assumptions of technological progress and positive ageing costs in a critical sense that give Turvey's model its true dynamic flavour - merely discounting the future is not enough to do this. This is exemplified by the number of so called dynamic neo-classical investment models such as those of Jorgenson(7) which turn out to be only a series of static decision problems. (A proof of these rather important observations is given in the Appendix, section (iv).)

Before we examine the rest of the model in detail we can ask what Turvey expects to get out of the model. In a real-world situation the solutions (Q^v) and (O_t^v) would be of over-riding importance, in an analytical exercise this is not the case and far more attention is directed to other products of the solution. Outstanding among these is a measure of marginal cost. It has long been known that in a cost minimisation problem marginal cost can be calculated as the effect on the objective function of a small change in the constraints; this is what Turvey is particularly interested in, and he is interested in it as a basis for developing pricing policy from a welfare economics point of view. Marginal cost therefore is exclusively a result of optimising behaviour for Turvey. In discussing this point he recognises (1969, page 282) that marginal cost may be calculated for other reasons but the optimising one is the main one for him. An important contribution which we will discuss below (Albouy and Nachtigal 1970) also recognises this:

> .... in static optimisation ... marginal cost remains an indispensable concept in interpreting the results.

In addition to calculating marginal cost Turvey is also interested in establishing socially optimal pricing and investment rules. He makes some headway with this especially in using his model to solve the peak load pricing problem (1971a Chapter 7; see also Littlechild 1970). One translation of these rules is to say that capacity is optimally adjusted to demand when short run marginal cost equals long run marginal cost. Turvey uses his model to develop this rule, however the outcome is not clearly resolved. One result of this was a severe criticism of

Turvey's 1969 paper by Kay (Kay 1971). In fact all of the discussion on this point appeared to be hopelessly mixed up and we can attempt to resolve it below(8). In the process we shall have something to say about the usefulness of short and long run concepts in dynamic optimisation. Finally Turvey develops from his model some interesting results on depreciation and amortisation - and, interestingly enough, some of these have appeared in different disguises in the earlier economic literature(9).

THE SOLUTION OF THE MODEL

We now have enough to start examining and solving the model in detail. Recalling the above discussion the objective is

> Minimise C = PW (total system costs from $t = o$ to $t = \infty$)
>
> by solving for the optimal capacity expansion programme
>
> $(Q^1 \quad Q^2 \quad \ldots\ldots \quad Q^v \quad \ldots\ldots\ldots)$
>
> and the optimal utilisation programme
>
> $(0^o_o \quad 0^o_1 \ldots 0^o_t \ldots 0^1_1 \ldots 0^1_t \ldots 0^v_t \ldots\ldots)$
>
> Where PW = the present worth at some specified discount rate of the stream of costs.

$$C = \sum_{v=o}^{\infty} \left[c^v Q^v + \sum_{t \geqslant v}^{\infty} r^v_t \; 0^v_t \right] \quad \ldots\ldots\ldots\ldots \; (1)$$

We note that one of the solution values is 0^o_o the rate of output in year o from capacity existing in year o; this is capacity inherited from the past, Q^o, since the first capacity solution value is Q^1, capacity to be built in the first year of the plan.

The most important constraint is that output rates are to meet the specified set of demand forecasts: total output from all plants in year t is to be at least as great as demand in year t :

$$\sum_{v \geq o}^{t} O_t^v \geq X_t \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

the remaining constraints relate to the operating technology of the system

constraint : Output may never exceed capacity for any machine

$$O_t^v \leq Q^v \quad \text{for all t and v} \qquad \ldots\ldots\ldots\ldots \text{(3)}$$

constraint : capacity available at the beginning of the plan cannot exceed the amount inherited from the past

$$Q^o \leq \bar{Q}^o \qquad \ldots\ldots\ldots\ldots \text{(4)}$$

constraint : output and capacity may never be negative

$$O_t^v \geq o; \quad Q^v \geq o \quad \text{for all t and v} \qquad \ldots\ldots\ldots\ldots \text{(5)}$$

There are no other constraints in Turvey's models as explicitly stated though he does indicate the possibility of making some; in particular, transmission constraints and possibilities which could make the system even more interdependent are ignored here.

We now have to solve the problem of minimising the objective function (1) subject to the constraints (2), (3), (4) and (5). The basis for the solution is known as the Kuhn Tucker theorem (see appendix); we can confine our attention to the important necessary conditions.

The first step is to create from (1) to (5) a single objective function without constraints. To do this we use a set of auxiliary variables called Kuhn Tucker multipliers : the amended objective function is therefore dependent not only on O_t^v and Q^v but also on these auxiliary variables.

m_t corresponding to the t constraints in equation (2)

k_t^v corresponding to the (vt) constraints in equation (3)

u_o^o corresponding to the single constraint in equation (4)

and these (like c^v and r_t^v) can be regarded as being in present value terms. The point of this is that each constraint reflects a condition that must hold at each point of time in the planning horizon and the strength of these constraints will diminish through time in our evaluation just as the costs of our planned operations c^v and r_t^v diminish through time. We can conceive of a current value Kuhn Tucker multiplier - say M_t - and its present worth $m_t = M_t\ (1 + i)^{-t}$ and we work in terms of m_t.

Our objective function is then:

$$L(0_t^v, Q^v, m_t, k_t^v, L_o^o) =$$

$$\sum_{v=o}^{\infty} \left[c^v Q^v + \sum_{t \geqslant v}^{\infty} r_t^v\ 0_t^v \right]$$

$$+ \sum_{t=o}^{\infty} m_t \left[X_t - \sum_{v=o}^{\infty} 0_t^v \right]$$

$$+ \sum_{v=o}^{\infty} \sum_{t \geqslant v}^{\infty} k_t^v \left[0_t^v - Q^v \right]$$

$$+ u_o^o \left[Q^o - \bar{Q}^o \right] \qquad \text{..........} \quad (6)$$

The Kuhn Tucker theorem now allows us to solve the original problem as follows: under certain general conditions a set of values of O_t^v Q^v m_t k_t^v and u_o^o that provide a minimum of L for O_t^v and Q_t^v and Q^v and a maximum of L for m_t k_t^v and u_o^o also provide a minimum of C for O_t^v and Q^v subject to the constraints specified above. We therefore only have to examine the necessary conditions for L to be a minimum for O_t^v Q^v and a maximum for m_t k_t^v and u_o^o.

Taking these in turn, we proceed as follows.

<u>Optimal choice of the use of inherited capacity Q^o</u> : if Q^o is to be optimal a small rise in it must either raise or leave constrained costs unchanged : therefore we require

$$\frac{\partial L}{\partial Q^o} \geqslant o \text{ and } Q^o \frac{\partial L}{\partial Q^o} = o$$

where $\partial L/\partial Q^o$ is the effect on constrained system costs of using some additional inherited capacity and where the second term expresses the fact that if a small rise in Q^o raises costs then we would prefer not to use any Q^o at all;

$$\text{and } \frac{\partial L}{\partial Q^o} = u_o^o - \sum_{t=o}^{\infty} k_t^o \qquad \ldots\ldots\ldots\ldots (7)$$

<u>Optimal installation of new capacity Q^v</u> : our reasoning is similar - we install Q^v up to the point at which any further installation raises total system costs - if this arises for any level of Q^v then we install no capacity of that vintage.

$$\frac{\partial L}{\partial Q^v} = c^v - \sum_{t\ v}^{\infty} k_t^v \geqslant o \text{ and } Q^v \frac{\partial L}{\partial Q^v} = o \text{ for all v} \qquad \ldots\ldots\ldots\ldots (8)$$

where $\partial L/\partial Q^v$ is the effect on system costs of installing additional Q^v capacity.

<u>Optimal choice of output rates O_t^v</u> : Proceeding similarly we have positive output from a given Q^v in year t if to do so does not raise costs; otherwise we have zero output from that plant in year t.

$$\frac{\partial L}{\partial O_t^v} = r_t^v - m_t + k_t^r \geq o \text{ and } O_t^v \frac{\partial L}{\partial O_t^v} = o \text{ for all v and t.} \quad \ldots\ldots\ldots\ldots\ (9)$$

Our arguments with effect to the constraints are slightly different. For each constraint we must judge whether - at the optimum - that constraint is an effective limitation on our freedom of action. If it is then the effect of relaxing it a little must be to reduce costs. Reasoning this way we obtain as necessary conditions for constraints on total output:

$$\frac{\partial L}{\partial m_t} = (X_t - \sum_{v=o}^{t} O_t^v) \leq o \text{ and } m_t \frac{\partial L}{\partial m_t} = o \text{ for all t} \ \ldots\ldots\ (10)$$

Where $\partial L/\partial m_t$ is the effect on system costs of a small rise in the "shadow cost" arising out of the existence of this constraint, and

$$\frac{\partial L}{\partial k_t^v} = (O_t^v - Q^v) \leq o \text{ and } k_t^v \frac{\partial L}{\partial k_t^v} = o \text{ for all v and t} \ \ldots\ (11)$$

and

$$\frac{\partial L}{\partial u_o^o} = (Q^o - \bar{Q}^o) \leq o \text{ and } u_o^o \frac{\partial L}{\partial u_o^o} = o \quad \ldots\ldots\ldots\ (12)$$

These necessary conditions ensure that where a constraint is effective the optimal values are indeed feasible values. Finally the Kuhn Tucker theorem ensures that O_t^v Q^v m_t k_t^v and u_o^o chosen in this way will never be negative (our constraint equation (5) is satisfied automatically).

Having solved his model in this way, Turvey proceeds to use it to obtain a variety of important results. The model is very flexible and rich and as well as examining some of Turvey's chief propositions we can proceed ourselves to discover some of the nuances of the model that Turvey himself does not mention.

Of course in any practical use of this model (or ones like it) attention is focussed on the optimal 'paths' of capacity installation and utilization (Q^v) and O_t^v); however in an abstract study we concentrate on analytical conclusions. A good starting point is equation (9). Essentially whenever a machine produces a positive rate of output in a given year : $O_t^v > o$, we know that, at the optimum a small rise in the rate produces no rise in costs:

$$r_t^v - m_t + k_t^v = o$$

so that

$$m_t = r_t^v + k_t^v \qquad \text{.......... (13)}$$

m_t only varies with t so that there is, in any one year, a set of implied equalities over different vintages of plant:

$$r_t^o + k_t^o = r_t^1 + k_t^1 = \ldots = r_t^v + k_t^v = \ldots\ r_t^t + k_t^t = m_t \qquad \text{........ (14)}$$

We now need to interpret k_t^v and m_t to understand this proposition. k_t^v and m_t are Kuhn Tucker multipliers or "dual variables" and are crucial to analysis of the underlying structure of the problem. Their strategic importance is expressed by showing that - under special conditions:

the value of a Kuhn Tucker multiplier (or dual variable) at the optimum = the effect of the optimal value of the objective function of a small relaxation in the constraint associated with that multiplier.

For example if the special conditions were satisfied we could write,

$$m_t = \frac{\Delta C^*}{\Delta X_t} \quad \text{at the optimum value of } m_t$$

Where C* is the value of system costs at the minimum point and C* is a function of the optimally chosen variables O_t^v Q^v which are in turn functions of the exogenously given constants. Turvey does not investigate these special conditions and they are very rarely treated in the literature so it is worthwhile spending a few moments examining them. Hence the next section is a digression on this topic. The main flow of the argument is taken up once more beyond this section.

THE SHADOW PRICE INTERPRETATION OF KUHN TUCKER MULTIPLIERS

In this section we have digressed to consider the conditions under which it can be said that

$$m_t = \frac{\Delta C^*}{\Delta X_t}$$

This is developed from the process of relaxing the appropriate constraint in the problem. The exogenous variables of the problem are X_t, r_t^v and c^v and the cost minimising solution C^* depends on the solution values O_t^v, Q^v and these in turn depend on X_t, r_t^v and c^v. If we were now to change X_t, say, by a small amount, some values of O_t^v and Q^v would change and hence C^* would also change. If we could evaluate these changes over a small range we can measure - in this example

$$\frac{\Delta C^*}{\Delta X_t} = \frac{\Delta C^*}{\partial O_t^v} \cdot \frac{\Delta O_t^v}{\partial X_t} \quad \ldots \quad \text{etc.}$$

However we must ask the question whether this is always possible. Is it the case that a finite

$$\frac{\Delta C^*}{\Delta X_t}$$

exists for every change in X_t? Once we have tackled this, there is then the further question of whether

$$\frac{\Delta C^*}{\Delta X_t} = m_t$$

Tackling the first problem, we want to consider whether a finite $\Delta C^*/\Delta X_t$ exists. Precisely this problem was tackled by Balinski and Baumol (1968), and the essential points can be conveyed diagrammatically. Suppose all our variables are reduced to two, labelled Y_1 and Y_2.

We are dealing with a minimum cost problem so the structure is simply

$$\text{minimize } C = F(Y_1\ Y_2)$$

subject to

$$g_1\ (Y_1\ Y_2) \geqslant X_1$$

$$g_2\ (Y_1\ Y_2) \geqslant X_2$$

etc.

$$Y_1 \geqslant o, \quad Y_2 \geqslant o, \quad X_1 \geqslant o, \ X_2 \geqslant o \quad \ldots. \text{ etc.}$$

If this was a linear problem (linear objective function and linear constraints) - and of course Turvey's problem is entirely linear - we could picture it as in figure VI,2 where the shaded area is the feasible region initially. (This figure is based on the analysis of Balinski and Baumol, particularly their figure 1a but adapted to deal with a cost minimizing formulation.) The constraints are "greater than" inequalities and are the lines AA^1, BB^1, CC^1. Two isocost contours of the objective function are shown for completeness. Initially a solution will occur at point B where an isocost contour reaches the lowest feasible level. Suppose now that the constraint equation represented by BB^1 is relaxed, so that the appropriate constraint is $\hat{B}\hat{B}^1$ with the boundary of the feasible region being ADEC. The new solution will be given by isocost contour passing through D but the effect on cost is simply to lower it in proportion to the relaxation of the constraint: cost falls in proportion to the distance BD. But this fall in cost may not always occur so smoothly. For example suppose there was a further relaxation in this constraint that took it down to the level $\tilde{B}\tilde{B}^1$. The feasible region however becomes AQC and the relaxed constraint is no longer binding: abruptly $\Delta C^*/\Delta X$ has become zero. Such discontinuities might also occur if the constraint remained binding but the optimal solution shifted to a new vector. If we graph C* as a function of X_1, the exogenous variable,

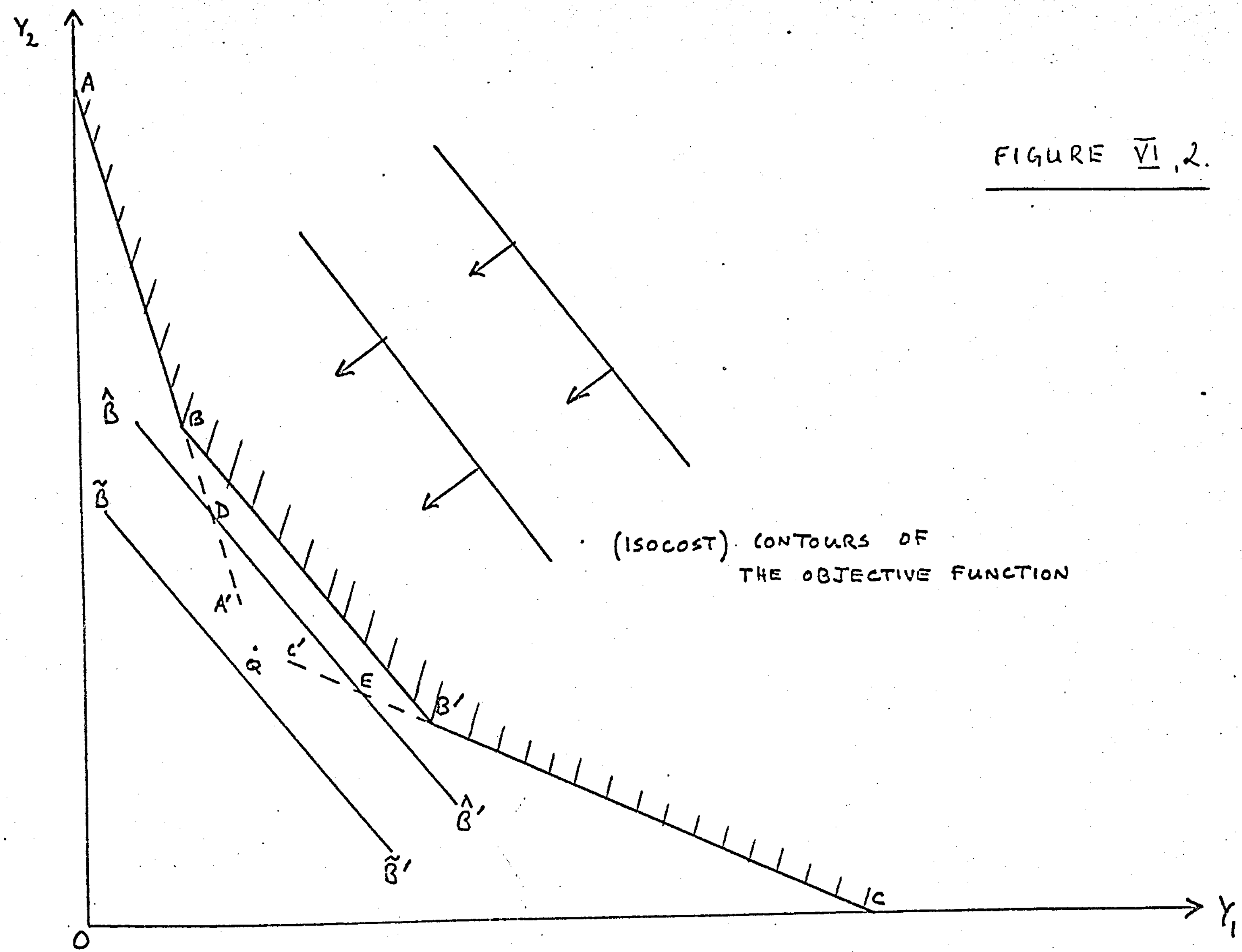


FIGURE VI.2.

140(a)

we therefore get a result like figure VI,3. For values of X less than OX^1, the derivative

$$\frac{\Delta C^*}{\Delta X}$$

is well defined. But for OX^1 (and also OX^{11}) $\Delta C^*/\Delta X$ is not defined.

At points like OX^1 and OX^{11}, small changes in X so radically alter the structure of the optimal plan that C* does not change in a smooth manner at all. This appears to raise the theoretical problem that marginal cost for some levels of output may simply not be measurable.

However Balinski and Baumol did establish a saving result, since they were able to show that

$$\Delta C^*/\Delta X$$

for OX^1 and similar points would lie between the finite limits given by the values of $\Delta C^*/\Delta X$ to the left and right of OX^1. This means that $\Delta C^*/\Delta X$ at OX^1 would be given by the slope of a line tangential to the C*(X) curve passing through W and lying between OW and WV. Marginal cost could therefore be measured within limits. (The essential condition is that the constraint qualification necessary for the Kuhn Tucker solution should hold; see appendix.)

Hence the answer to our first problem is as follows. Given the conditions for a solution to the non-linear programming problem to exist (the Kuhn Tucker constraint qualification), then relaxing a constraint produces a measure

$$\frac{\Delta C^*}{\Delta X_t}$$

which at worst will be defined within positive limits.

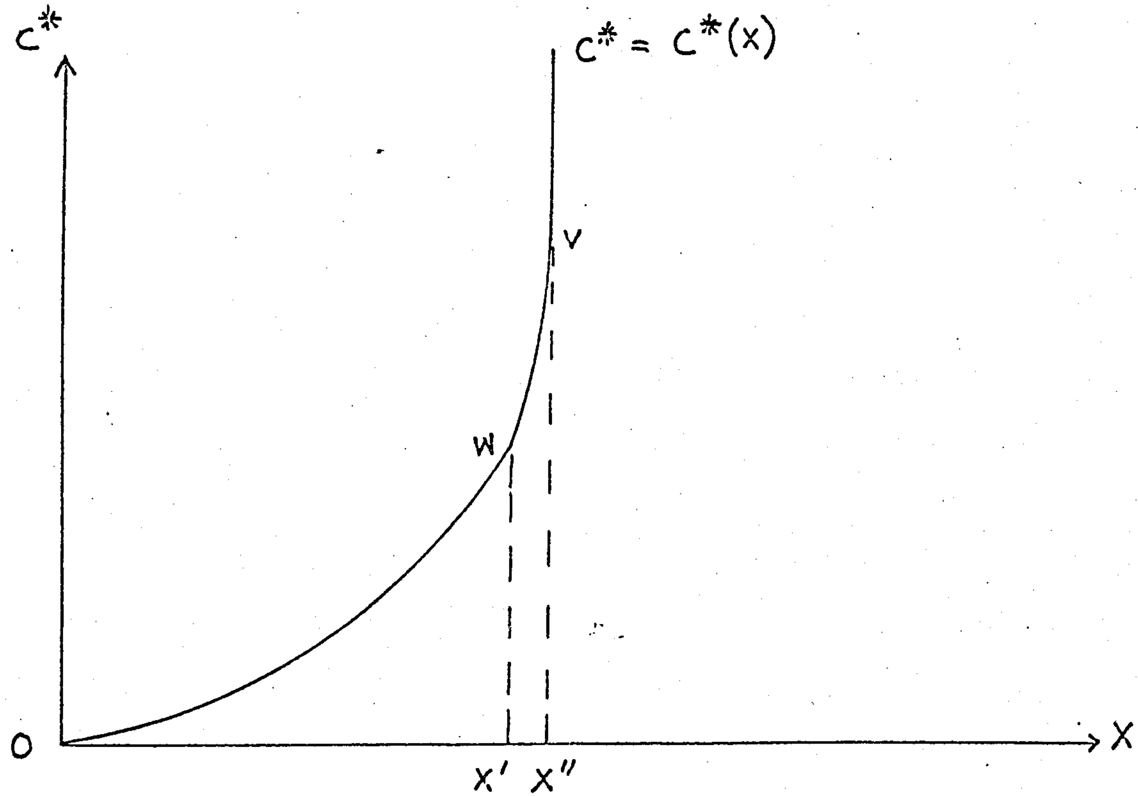


FIGURE VI, 3

As we have already remarked, Turvey places weight on this measure of marginal cost in pricing policy, so it has to be accepted that for some output ranges the nearest we may get to setting a price based on marginal cost is to pick a price from a finite range of values.

The second problem we pointed out was this:

given that we know $\Delta C^*/\Delta X_t$ exists within finite limits, is it true that it can be measured by m_t, the Kuhn Tucker multiplier associated with the constraint containing X_t?

This is simply a matter now of careful analysis of the optimum solution.

Recalling the optimal solution we know that the optimum value of L depends on Q^v, 0_t^v, m_t, k_t^v, u_o^o

$$L^* = L^* \; (Q^v \;\; 0_t^v \;\; m_t \;\; k_t^v \;\; u_o^o\,) \text{ at the optimum}$$

But then allowing for these variables in turn to be affected by $X_t \; r_t^v \;\; c^v$ we ought to write

$$L^* = L^* \left[Q^v \; (X_t \; r_t^v \; c^v) \; 0_t^v \; (X_t \; r_t^v \; c^v),\; m_t \; (X_t \; r_t^v \; c^v) \; k_t^v \; (X_t \; r_t^v \; c^v) \; u_o^o \; (X_t \; r_t^v \; c^v) \right]$$

We are particularly interested here in X_t holding c^v and r_t^v constant. Let us therefore focus our attention on

$$L^* = L^* \left[Q^v \; (X_t) \; 0_t^v \; (X_t) \; m_t \; (X_t) \;\; k_t^v \;\; (X_t) \;\; u_o^o \;\; (X_t) \right]$$

Where the optimal values of the variables all in turn depend on the given demand forecasts X_t. We are now able to examine a small change in X_t

by investigating what happens to these optimal values as X_t changes

$$\frac{\Delta L^*}{\Delta X_t} = \sum_v \frac{\partial L^*}{\partial Q^v} \frac{\Delta Q^v}{\Delta X_t} + \sum_v \sum_t \frac{\partial L^*}{\partial O_t^v} \frac{\Delta O_t^v}{\Delta X_t} + \sum_t \frac{\partial L^*}{\partial m_t} \frac{\Delta m_t}{\Delta X_t}$$

$$+ \sum_v \sum_t \frac{\partial L^*}{\partial k_t^v} \frac{\Delta \dot{k}_t^v}{\Delta X_t} + u_o^o \frac{\partial L^*}{\partial u_o^o} \frac{\Delta u_o^o}{\Delta X_t} + m_t$$

We can see that $\Delta L^*/\Delta X_t$ contains a series of complex terms and the single term m_t. This single expression m_t arises because the constraint containing X_t enters the expression for L^* by being multiplied by m_t. Hence any effect of ΔX_t on L^* will include at least this term reflecting the shadow price of this constraint; the question remains whether any of the other terms also enter. In a linear or non-linear programming framework we can never be quite sure whether terms like $\partial L^*/\partial O_t^v$ will be negative or zero or whether terms like $\partial L^*/\partial k_t^v$ will be positive or zero since either may be the case at or near the boundaries of the problem.

To keep matters simple suppose almost every term except m_t _is_ zero so that we are only left with the case of

$$\frac{\Delta L^*}{\Delta X_t} = m_t + \frac{\partial L^*}{\partial O_t^v} \frac{\Delta O_t^v}{\Delta X_t}$$

Which simply says that the cost of changing X_t by ΔX_t amounts to the shadow price, m_t on the X_t constraint and an item representing the cost of using a little more output O_t^v, from capacity of vintage v in period t. What can we say about this term bearing in mind that we only know that $\partial L^*/\partial O_t^v$ is not negative?

We reason as follows. Firstly, suppose $\partial L^*/\partial O_t^v$ is positive - any rise in O_t^v raises costs; in this case we would not want to meet X_t at the optimum by using O_t^v since our costs will rise away from the optimum. But if this is the case $\Delta O_t^v/\Delta X_t$ is zero and $\Delta L^*/\Delta X_t = m_t$. Suppose now that $\Delta O_t^v/\Delta X_t$ is not zero - so that near the optimum a variation in X_t will cause us to vary O_t^v. This might be the case if O_t^v is already part of our optimal solution. But in that case $\partial L^*/\partial O_t^v = o$ and again $\Delta L^*/\Delta X_t = m_t$. Finally suppose an infinitesimally small rise in X_t caused us to raise O_t^v from zero to a small positive amount. In that case O_t^v though not a part of the optimal solution must have been on the verge of being a candidate for the optimal solution. Then $\partial L^*/\partial O_t^v$ must have been almost zero. Hence we can show that in the neighbourhood of the optimum

either $\Delta O_t^v/\Delta X_t$ is very close to zero

or $\partial L^*/\partial O_t^v$ is very close to zero

and therefore write

$$\frac{\Delta L^*}{\Delta X_t} = m_t$$

Our final step is then to note that at the optimum $\Delta L^*/\Delta X_t$, the effect on the optimal value of our constrained costs, is exactly equal to $\Delta C^*/\Delta X_t$ the effect on the optimal value of unconstrained costs. (See appendix on Kuhn Tucker theorem).

Hence

$$\Delta C^*/\Delta X_t = m_t$$

We have established two things in this section for the case of non-linear programming problems:

(a) At the optimum $\Delta C^*/\Delta X_t$ exists and is finite but may be discontinuous, if the constraint associated with m_t is binding.

(b) At the optimum $\Delta C^*/\Delta X_t$ must be very close to m_t.

APPLICATIONS TO TURVEY'S MODEL We can now continue our interpretation of equation (14) i.e.

$$r_t^o + k_t^o = r_t^1 + k_t^1 = \ldots r_t^v + k_t^v = \ldots . r_t^t + k_t^t = m_t \quad \ldots . (14)$$

We can therefore interpret $m_t = \Delta C^*/\Delta X_t$ in two ways. One is as a marginal cost concept of the forecast demand levels, X_t; the other is in terms of the optimal utilisation of an item of capacity. m_t is the sum of running cost, r_t^v and a dual variable or shadow price k_t^v for each machine. The two together make up a given m_t for any year. Clearly therefore at the optimum the planner in meeting any forecast ΔX_t is indifferent among a whole complex of plants ranging from the newest installed to the oldest in operation.

With our dynamic hypothesis about r_t^v we therefore obtain an idea of how k_t^v might look. Turvey has modified a celebrated diagram of Salter's to illustrate this result. Figure VI,4a and VI,4b show for a given demand level at the optimum, X_3^*, the relation between

$$\begin{array}{cccc} r_3^o & r_3^1 & r_3^2 & r_3^3 \\ k_3^o & k_3^1 & k_3^2 & k_3^3 \\ 0_3^o & 0_3^1 & 0_3^2 & 0_3^3 \end{array}$$

and m3.

FIGURE VI, 4a.

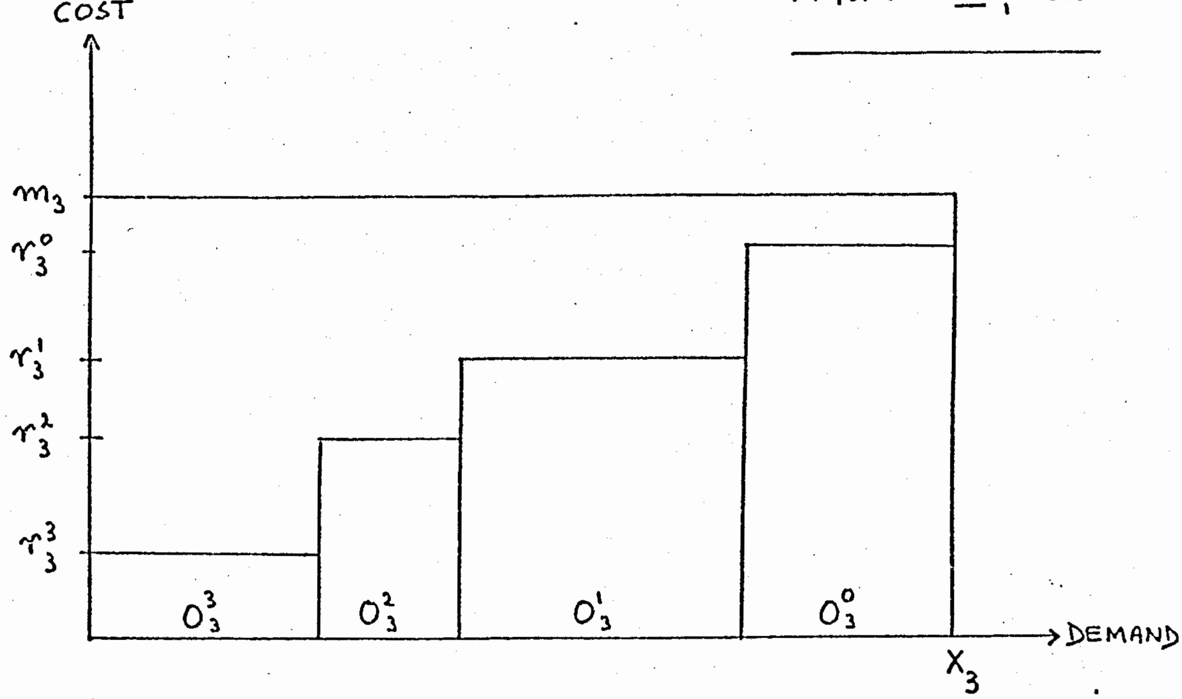
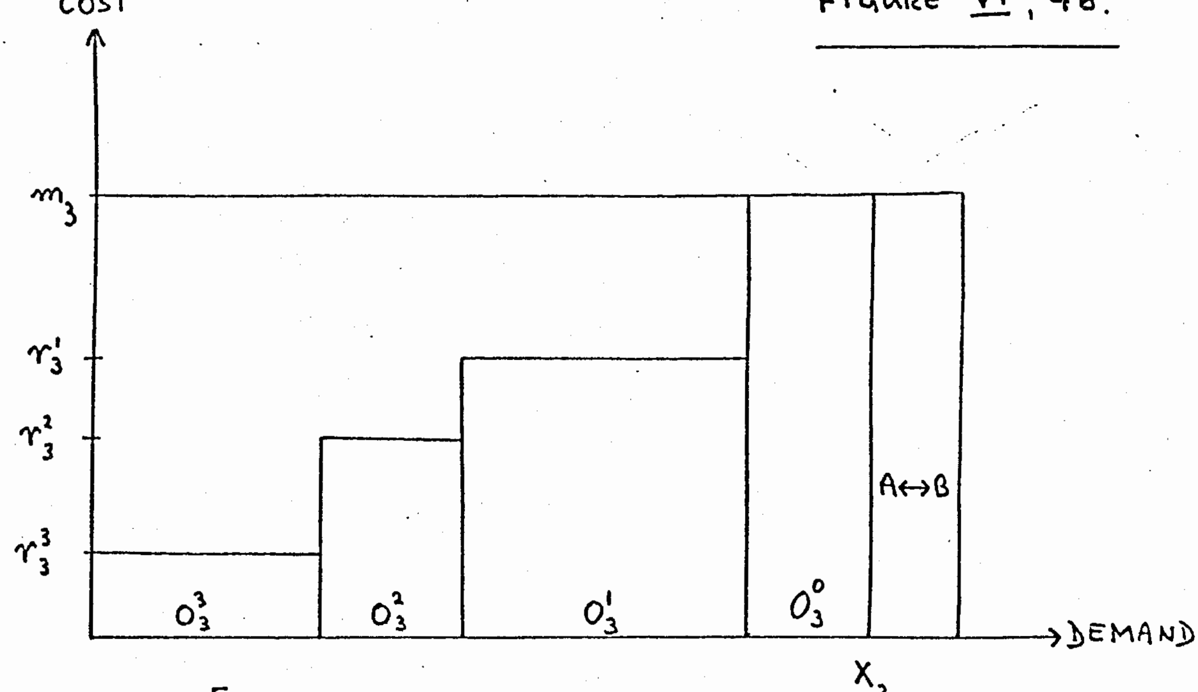


FIGURE VI, 4b.



[AB MEASURES THE UNUSED OUTPUT CAPACITY OF Q^0 IN YEAR 3.]

In figure VI,4a all plants are used to capacity

$$O_3^v = Q^v \text{ for all vintages } 0\ 1\ \ldots\ 3$$

While in figure VI,4b there is some capacity space on earliest vintage plant, (here Q^o) in period 3. In this later case it is a necessary part of our optimum condition that $k_3^o = o$ so that $m_3 = r_3^o$ and only running cost enters marginal cost for this marginally used plant.

Just as m_t is interpreted by Turvey so k_t^v provides a useful interpretation point. Turvey argues that

$$k_t^v = \frac{\Delta C^*}{\Delta Q^v}$$

the effect on system costs of having some additional capacity. Intuitively it is easily seen that it is the k_t^v term that brings running costs up to marginal cost for any machine in any period; k_t^v therefore is a capital charge associated with a particular machine built at t = v and operating in year t. This is a sensible interpretation though it is not quite the same as

$$\frac{\Delta C^*}{\Delta Q^v}$$

We already know at the optimum that

$$\frac{\Delta C^*}{\Delta Q^v} = \frac{\Delta L^*}{\Delta Q^v} \geq o$$

but Turvey, in calling k_t^v a cost <u>saving</u> associated with additional capacity (or quasi rent) is implying that k_t^v is negative.

It is more correct to say

(a) k_t^v is a form of capital charge

(b) it is also the effect on L* of slightly relaxing the constraint.

$$O_t^v - Q^v \leqslant o$$

Thus suppose the constraint is binding so that

$$O_t^v - Q^v = o$$

If the constraint was then written as

$$O_t^v - Q^v = e$$

Where e is a small positive number we would effectively have raised the available amount of Q^v and could then interpret as

$$k_t^v = \frac{\Delta L^*}{e}$$

It is necessary therefore to distinguish between the positive:

k_t^v a capital charge, and the idea of using k_t^v as a quasi-rent from Q^v in year t as Turvey does. In a cost minimising exercise quasi-rents have negative signs. Turvey in fact is stretching the economic ideas of this model over into a different type of model. As we shall show below(10), this second model was first developed by Hotelling (1925) and is an interesting antecedent of the present analysis.

Bearing this in mind, we can see that for each plant there is a running cost and capital charge item. For any plant in operation these add to marginal cost, m_t. What happens when a plant is marginal to the operating programme? As a plant reaches the end of its economic life its quasi-rents or cost savings tail off; at the optimum there will be zero capacity charges recoverable from the last period of its use

$$\text{i.e. if } O_t^v \leqslant Q^v, \quad k_t^v = o$$

This can be seen in figure VI,4b. Part of X_3 is made up by operating some old inherited plant but at a level below capacity. AB in figure VI,4b measures unused capacity of Q^o in year 3. The plant is in operation but its running costs are equal to the whole of marginal costs and it is on the verge of not being used at all.

$$m_3 = r_3^o$$

We have here a condition for use or non-use of capacity but not enough to give us an indication of whether that capacity ought to be scrapped. To know this we must not only look at k_3^o which is equal to zero but also at k_4^o k_5^o ... and so on. The factors making up our scrapping criterion are simply whether

$m_t \leqslant r_t^v + k_t^v$ for all t from now to infinity. The following chapter makes some additional comments on this proposition.

In developing the optimal planned expansion k_t^v not only plays a part in determining usage or scrapping of plant but also in acquisition.

Examine equation (8)

$$\frac{\partial L^*}{\partial Q^v} = - \sum_t k_t^v + c^v \geqslant o \; ; \quad Q^v \frac{\partial L^*}{\partial Q^v} = o$$

From this, if $Q^v > o$ then $\partial L^*/\partial Q^v = o$; i.e. a necessary condition for building capacity Q^v is that

$$c^v = \sum_t^{\infty} k_t^v$$

or, that the capacity cost c^v is just covered, for each item of capacity, by the lifetime allocation of capital charges associated with that capacity. Some of the k_t^v will be zero : this is immaterial - as long as c^v can be covered by allocating a part of it every time capacity is to be used then that capacity ought to be built. These two results:

$$m_t = r_t^v + k_t^v$$

$$c^v = \sum_t k_t^v$$ make up the essence of Turvey's model.

Marginal cost in any one year is given by a running charge and a capital charge for whatever plant is utilised; this can be called a short term pricing rule if m_t is to be used as a basis for pricing. Capacity cost is covered by a time stream of capital charge allocations for whatever capacity is built; this can be called an investment rule. Together the pricing rule and the investment rule determine optimal capacity expansion and utilisation.

There is much else in Turvey's model but one of the most interesting aspects still to be covered is what the model tells us about measuring m_t and k_t^v.

We can approach this by an indirect route. Let us look at equation (12)

$$\frac{\partial L^*}{\partial u_o^o} \leqslant o \quad \text{and} \quad u_o^o \frac{\partial L^*}{\partial u_o^o} = o \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

where $$\frac{\partial L^*}{\partial u_o^o} = -\sum_t k_t^o + u_o^o$$

Let us suppose $$u_o^o = \sum_t k_t^o$$

We must distinguish carefully between $\sum_t k_t^o$ and

$\sum_{\substack{t \geq v \\ v \geq o}} k_t^v$. The second expression is the present worth of the stream of capital charges from capacity which will be acquired as part of the optimal programme and, as such, it not only covers c^V but amounts to all the possible capital charges attributable to this capacity. But capacity Q^O is inherited from the past. It is not optimally chosen by our programme, it is simply given to us. As such we have no way of knowing if its capacity cost has been or will be covered. We can derive a shadow price for this capacity, u_o^o

$$u_o^o = \frac{\Delta C^*}{\Delta Q^o}$$

but this shadow price is simply a reflection of the remaining capital charges for Q^O - not the entire lifetime amount of its capital charges, some of which may have been allocated before our programme began.

Since we know $k_t^o = (m_t - r_t^o)$ from equation (14)

write $$u_o^o = \sum_{t \geq o} (m_t - r_t^o)$$

and u_o^o is the stream of cost savings <u>remaining</u> to some capacity inherited from the past. It could be called the "<u>unamortised</u>" value of inherited capacity.

Now we could imagine time marching on and our optimal plan becoming an actual programme passed on to future generations. Associated with each succeeding plan there would be a series of terms

$$u_t^v = \sum_{t \geq v} (m_t - r_t^v) \qquad \text{.......... (15)}$$

being the present worth of the remaining cost savings of capacity actually built. Of course for capacity about to be built

$$c^t = \sum_{n \geqslant t} (m_t - r_t^m) = u_t^n$$

the unamortised value of capacity Q^t is its total capacity cost - this is a special case of our general equation (15). Equation (15) can always be written:

$$u_t^v = k_v^v + \sum_{t \geqslant v} k_t^v = m_v - r_v^v + \sum_{t \geqslant v} (m_t - r_t^v)$$

so that

$$m_v = r_v^v + \left[u_t^v - \sum_{t \geqslant v} (m_t - r_t^v) \right] \qquad \text{..........} \quad (16)$$

The second term in (16) is of course k_t^v but more usefully is the fall in the unamortised value of Q^v over the next year's use - what many economists have called true depreciation. Let us relate this to equation (14)

$$m_t = r_t^o + \left[u_t^o - \sum_{n \geqslant t} (m_n - r_n^o) \right] = \dots = r_t^t + \left[u_t^t - \sum_{n \geqslant t} (m_n - r_n^t) \right]$$

and $u_t^t = c^t$

Marginal cost in a given year is the sum of running cost and true depreciation on each plant, and in particular the sum of running cost and first year amortisation on newest (current vintage) plant.

System interdependence is clearly emphasised by this result:

m_t depends on every subsequent m_n, $n \geqslant t$ (though not on previous m_s, $s < t$). The only way marginal cost can be calculated is by looking <u>forward</u> to expectations - it cannot be calculated by looking backwards. Turvey sums this up as follows:

> "... it is clear that first year amortisation epitomises the complex of expectations and calculations about the future which are central to the notion of marginal cost ... in principle an intelligent guess at first year amortisation could furnish an intelligent guess at marginal cost ...."
>
> (Turvey 1969 p.296)

This completes our exposition of the global programming model of Turvey in terms of its properties. Before turning to a consideration of its implications and its place in the other investment planning models of the literature we can summarise those aspects of it on which we have focussed attention.

(a) The model chooses an optimal path or sequence of capacity construction (Q^v) and utilisation (O^v_t) to minimise the present worth of the system costs of meeting some specified demand forecasts.

(b) The model emphasises two important characteristics of the investment planning problem: (1) system interdependence and (2) non stationarity; the first characteristic shows up in the importance that expected costs of future operations with items of capacity have on the costs of present operations with other items of capacity; the second characteristic - inherent in the assumptions about exogenous factor prices r^v_t and c^v - appears in the way the model determines the course of m_t and k^v_t. The optimum solutions remain inherently dynamic and dependent on the future.

(c) An interesting marginal cost concept $\Delta C^*/\Delta X_t$ is thrown up in the programming formulation. We know something about its discontinuities and that it is equal to one of the dual variables m_t. It remains to be seen what sort of marginal cost concept this is and what use Turvey makes of it.

(d) The model produces a pricing (or costing) rule and an investment rule which in turn tell us a great deal about how we might determine the time profile of amortisation on capacity.

NOTES TO CHAPTER VI

1. The classic work is Lipsey and Lancaster (1956). One aspect of the topic has appeared in part I's treatment of the way prices are related to marginal costs when a surplus is to be made. A general guide is Rees (1968).

2. Turvey 1971a page 54.

3. While this is true of the models examined in this chapter, nevertheless Chapter VIII does present an optimal investment planning model that explicitly incorporates a "load curve" into the analysis. In practical applications this latter approach may be computationally more useful.

4. The appendix referred to is the mathematical appendix to part II of the thesis.

5. The idea has been discussed in the optimal growth literature, most of which stems from the work of Ramsey (1928).

6. The references are Salter (1966), Solow (1959), Johansen (1959).

7. Jorgenson (1965) was the first of these neoclassical investment models.

8. The resolution appears in Chapter IX.

9. These are discussed at length in Chapter VII following.

10. See Chapter VII following.

CHAPTER VII

The Background to Turvey's Model of the Investment Planning Problem

Introduction

We now have a detailed knowledge of Turvey's programming model in its 1971a version, and it would be helpful at this point to examine some of the background to his propositions. Turvey has used and analysed his model on several occasions and these, especially the discussion in his 1969 paper on Marginal Cost (Turvey 1969) give some idea of the influences he has attempted to take into account. To begin with we can recall the nature of the basic problem. The planner wishes to find two optimal paths:

- the path of capacity accumulation or expansion
- the path of output - i.e. capacity utilization.

In doing this he faces a given objective function - in our case minimising total systems costs and he also must take into account constraints on his actions. The nature of the constraints in Turvey's model can be summed up as

- output should never exceed capacity
- output should never fall below the specified demand requirements.

The two crucial characteristics of the Turvey model of this situation are then

- system interdependence
- the non stationary character of input prices.

Not all of the literature has specified the problem in this way, and different models have focussed on different aspects of the optimal solution. With this in mind let us examine some of these other models.

OTHER MODELS OF THE INVESTMENT PLANNING PROBLEM

1. HOTELLING (1925) Of the previous attempts to tackle investment planning problems in the Turvey spirit one of the earliest and most penetrating was a model developed as early as 1925 by Harold Hotelling. Hotelling's work grew out of his dissatisfaction with the established methods of depreciating assets. He began by asking the fundamental question: What must be the "theoretical selling price" of the product of a machine before we can decide whether the machine should be acquired. Clearly it is meaningless to answer this question by adding running costs and "depreciation" when depreciation is defined as the rate of change of the value of the machine. The established models of the accounting literature did just this and were therefore irrelevant. A second important consideration for Hotelling was - once a machine was acquired - when ought the machine to be scrapped?

Hotelling sets up his model as follows. The fundamental variable is the annual rental value from the machine

$$K(t) = pO(t) - R(t)$$

p is the theoretical selling price described above, $O(t)$ is the rate of output of the machine and $R(t)$ is its running costs. $O(t)$ is specified in the blueprint for the machine as its maximum output capacity. Of course a planner could decide to run the machine below capacity, but Hotelling leaves aside this possibility in the initial part of his analysis. $O(t)$ and $R(t)$ are therefore known time paths. Once a machine has been acquired at time $t = o$, its value from then until the end of its life depends simply on p and n where n is the year in which it is scrapped - given that it is always run at the fixed maximum rate of output O(t). Thus suppose we have acquired a machine previously and now (time $t = \tau$) are looking at its future rentals, the value remaining in the machine is

present worth from $t = \tau$ to $t=n$ (annual rentals from τ to n) + present worth in year τ (scrap value when machine is scrapped in year n)

Scrap value is the known value of a machine aged n years on the second hand market. Hotelling assumes this market is perfect and scrap value of machinery is identical to the "inside value" of the machine to the owner - the value remaining in the machine in its place in the owner's plant. If we use continuous discounting with δ as our discount rate we can write down the above ideas as

$$V(t) = \int_{\tau}^{n} \left[p\,O(t) - R(t) \right] e^{-\delta(t - \tau)}\,dt + S(n)e^{-\delta(n - \tau)} \quad \ldots\ldots\ldots (1)$$

So much for the value remaining in a previously acquired machine. But when should a machine be acquired? Clearly when:

Cost of acquiring machine (C) = total value of machine over its entire life $t = o$ to $t = n$ in present worth terms.

i.e.

$$C = V(0) = \int_{o}^{n} (pO(t) - R(t))\,e^{-\delta t}\,dt + S(n)\,e^{-\delta n} \quad \ldots\ldots\ldots (2)$$

The planners problem then is (a) whether to acquire a machine; (b) once acquired to scrap it in a way to maximize V(t).

Take the second problem first. The machine has been acquired because (p) the crucial price variable was known to be sufficient to warrant acquisition. When should the machine be scrapped? We have to picture what happens to $K(t) = pO(t) - R(t)$ over the machine's life. Two examples are given in figures VII,1a and VII,1b. Figure VII,1a shows a straightforward case of the machine giving declining output with rising running costs - the sort of picture we have in general of most capital equipment. Figure VII,1b shows the same fundamental idea but a much more complex pattern. On three occasions quasi rent $K(t)$ is zero. To begin with the machine takes some initial period to reach a high output and it experiences a new lease of life towards the end of its existence when its quasi rent almost becomes positive once more.

Hotelling labels these crucial time periods when output ceases to be worth operating expenses as m: m is given by

$$pO(m) - R(m) = o$$

In figure VII,1a there is only one value for m but in figure VII,1b there are 3 values for m and we would wish to choose the value of m which makes $V(t)$ a maximum in accordance with our overall objective. In figure VII,1b this will almost certainly be $m = m_2$. Now fixing on the appropriate value of m the sum of all the positive rentals from the machine will be

present worth from $t = \tau$ to $t = m$ of $K(t)$

$$= \int_{\tau}^{m} K(t) \; e^{-\delta (t - \tau)} dt$$

But before the period m is reached this sum of positive rentals remaining as we look at different years τ falls until it will coincide - in year n - with the scrap value of the machine.

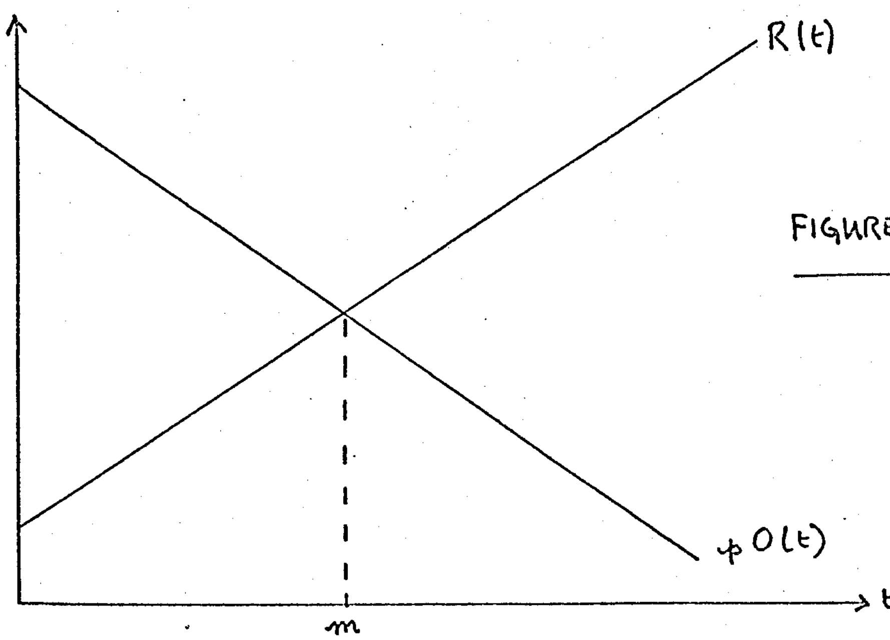


FIGURE $\overline{\text{VII}}$, 1a

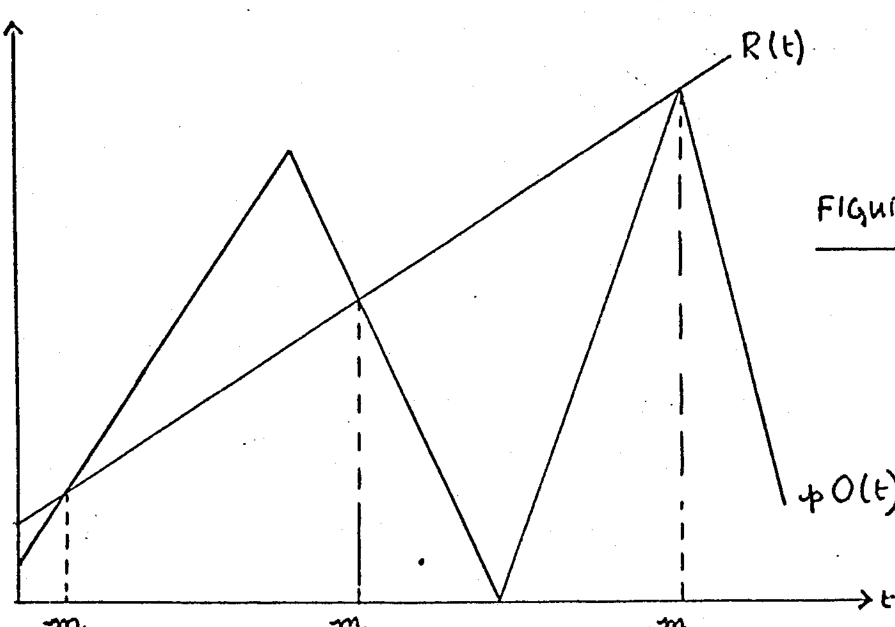


FIGURE $\overline{\text{VII}}$, 1b

HYPOTHETICAL PROFILES OF THE QUASI-RENTS FROM A UNIT OF CAPITAL EQUIPMENT : $K(t) = p\,O(t) - R(t)$

In year n

present worth of remaining positive rentals from n to m	$\equiv$	value of the machine in year n on the second hand market in present worth terms

i.e.

$$\int_n^m K(t)\, e^{-\delta(t-n)}\, dt = S(n) \qquad \ldots\ldots\ldots (3)$$

The owner is indifferent in year n between keeping and using his equipment and selling it for scrap, and in this way the year of scrapping or replacement is chosen. One aspect of Turvey's model which we did not investigate was whether or not it produced a general scrapping or replacement criterion like Hotelling's. In fact it is possible to derive from Turvey's model a replacement criterion analogous and rather similar to (3) but it is possible only with difficulty. This is because Turvey's model - although inherently dynamic because of the non stationary assumptions about factor prices - is nevertheless not solved sequentially but rather is solved simultaneously treating each value at each time period as a separate variable. To see this we can proceed as follows. In a solution to the Turvey model an item of capacity has been scrapped or replaced when it is apparent that it produces no more output; i.e. beyond a certain value of t, O_t^v equals zero for every point in time for certain vintages of capacity. The scrapping criterion in other words relates to the optimal values

$$O_t^v$$

and a machine is scrapped in year n if

$$O_t^v = o \text{ for all } t \geqslant n$$

The drawback of this form is that while the methods of optimization we have used do indicate which variables <u>necessarily</u> will have non zero values, they do not indicate which variables <u>necessarily</u> have zero values. Thus, capacity is utilized if doing so reduces costs

$$\frac{\partial L}{\partial O_t^v} > o \quad \text{and} \quad O_t^v \cdot \frac{\partial L}{\partial O_t^v} = o$$

but there is no way of deriving from these expressions, necessary conditions for $O_t^v = o$.

We can see that if $\partial L/\partial O_t^v > o$ then this implies that $O_t^v = o$ and so is sufficient; however if $O_t^v = o$ this does not imply that $\partial L/\partial O_t^v > o$ so the condition is not necessary. Concentrating therefore on sufficient conditions:

$\partial L/\partial O_t^v > o$ means that $r_t^v - m_t + k_t^v > o$

that is

$$k_t^v > m_t - r_t^v$$

and this condition must hold for all time periods, t, beyond the period n if the capacity Q^v is to be scrapped in period n. What does this condition mean? We know that for a feasible solution, with constraints binding, each of the terms above, k_t^v, m_t, r_t^v is positive. Now k_t^v is the potential additional cost from relaxing the constraint $O_t^v = Q^v$ by a small amount, say E; $(m_t - r_t^v)$ is the maximum additional cost that can be permitted to O_t^v for O_t^v still to play a part in the optimum; m_t of course is set by running and amortization costs on the <u>other</u> capacity within the system - hence the replacement rule in Turvey's model is:

replace capacity Q^V in year n if - for every time period beyond n - the additional cost associated with an extra portion of Q^V capacity is higher than the maximum additional cost: $(m_t - r_t^V)$ which would permit output from Q^V - i.e. O_t^V - still to be part of the optimum output path.

This in turn is sufficient to ensure that if Q^V is scrapped in year n, then

$$\sum_{t \geqslant n}^{\infty} k_t^V > \sum_{t \geqslant n}^{\infty} (m_t - r_t^V)$$

but this condition is <u>not necessary</u> for scrapping. This is about as far as we can go in Turvey's model towards obtaining a general scrapping rule; so on this account we have to conclude that the Hotelling model is partially superior to Turvey's formulation.

The more fundamental part of Hotelling's model, however, relates to the decision to acquire capacity and the determination of optimal pricing through time for the product of that capacity. Hotelling's main analysis continues to use the assumption that the machine will be needed to full capacity throughout its life. The problem then is to solve for p(t) which can be represented either as a contant or as a known function of time and an unknown parameter: e.g. $p(t) = r^t$ where r is some known figure. The two equations describing Hotelling's model above are then all that is needed. Equation (1) gives a scrapping condition by choosing n to maximise V(t); the condition $\frac{dV(t)}{dn} = o$ determines n and when this is inserted into equation (2), p(t) can also be determined. Hotelling, however, is conscious that this sort of approach masks the real nature of the problem and in the last part of his article concludes that the assumption that capacity is always fully used is not an essential part of investment planning analysis. But the dropping of this assumption completely changes the approach to the problem. Recall equation (1)

$$V(\tau) = \int_{\tau}^{n} \left[p(t)O(t)-R(t) \right] e^{-\delta(t - \tau)}dt+S(n)e^{-\delta(n - \tau)} \quad \ldots\ldots\ldots (1)$$

Now $V(\tau)$ can no longer be regarded merely as a function of p and n; $V(\tau)$ has become a functional: that is to say it will be satisfied only by choice of some optimal function, $O(t)$, describing for each instant in time the rate of output or capacity utilisation. The techniques for solving this type of problem are much more involved than the simple solution of simultaneous equations which represent optimality conditions that only depend on a few parameters. One method is to treat time as discrete and employ the Kuhn Tucker programming methods of Turvey; an alternative is to employ theorems on dynamic optimization treating time as a continuous variable. Hotelling recognised that this was at once a more difficult but more general approach than the usual rule of thumb depreciation methods.

2. WRIGHT (1964, 1968) The optimal investment planning model had been posed by Hotelling as part of his search for the ideal depreciation method. Hotelling, once he had shown how the problem of choosing a maximum of $V(\tau)$ should be solved, obtained optimal depreciation - the amortization of Turvey's model - as the rate of change of $V(\tau)$: $dV(t)/dt$. The question of optimal depreciation of long lived assets has continued to be the motivating source for treatments of optimal investment in the accounting literature. Wright's models - (particularly Wright 1968) - developing Hotelling's ideas end up with a formulation very close to the model of Turvey that we have examined in detail. Wright's aim is to develop a measure of what he calls opportunity value of an asset : the loss of value to an

enterprise associated with the disappearance of the services of a given asset at a particular period of time. The course of opportunity value over time is then taken as the theoretically sound way of depreciating or amortising an asset. (Wright 1964, page 277). Wright's results can be examined in terms of his most general model (Wright 1968, page 299). The objective is to maximize the present worth of profits which are simply the discounted stream of sales at a fixed price less operating costs and expenditure on new capacity. Two constraints operate on this objective.

(a) both inherited and newly purchased equipment has a fixed capacity which cannot be exceeded by output in any period.

(b) at the fixed set of prices a specified maximum level of output can be sold and production cannot exceed this in any period.

Wright is particularly interested in the dual variables of this problem. Bearing in mind the analysis of the previous chapter we know that the dual variables associated with constraint (a) (Wright's W_{jk}) represents the value to the firm of producing a little extra from some given amount of capacity (of type j in period k) while those associated with constraint (b) (his V_{Ik}) represented the value of being able to sell an extra unit of output (of product I in period k). His solution to the depreciation problem then focusses on the W_{jk} since one of the optimality conditions shows that for any machine that <u>is</u> purchased, the capital cost is exactly accounted for by the lifetime present worth of the values (W_{jk}) of producing a little more from that capacity, i.e. the present worth of the values of just relaxing the constraint relating output to capacity. This is identically the concept which Turvey tries to capture in his k_t^v, except that here the dual variable is in terms of cost savings since his objective function is to minimise costs whereas for Wright W_{jk} is in terms of profits since his objective function is to maximise profits over the firm's life. Both models, therefore, adopt very much the same approach to the problem as can be seen by the comparison:

	TURVEY	WRIGHT
objective:	minimise present worth of system costs	maximise present worth of system profits
constraints	(a) output does not exceed capacity	(a) output does not exceed capacity
	(b) output does not fall short of demand	(b) output does not fall short of demand
results	estimates of marginal cost $\Delta C^*/\Delta X_t$ and estimates of optimal amortization in terms of cost savings	estimates of marginal revenue and estimates of optimal amortization in terms of marginal opportunity values at fixed prices
assumptions:	system interdependence and variation through time of factor prices	system interdependence and variation through time of factor prices

Wright uses the assumptions of the model and the consequent results to negate two well established alternative amortisation methods:

(i) revenue matching: system interdependence, which we saw illustrated in terms of Turvey's model, makes it impossible to associate with one machine in one period a specific fraction of total profits and hence attempts to depreciate assets by writing off in each year a proportion of revenues for which they are responsible are meaningless

(ii) rules of thumb - like straight line or diminishing balance depreciation - are at best only usable as approximations to ideal amortisation though diminishing balance may be just preferable if it reflectsmore accurately the effect on amortisation of technical progress embodied in the latest vintages of capacity.

What we have shown here is that a line of models, beginning with Hotelling's early contribution, that have approached the problem of optimal depreciation produce virtually identical results - under the appropriate assumptions - to Turvey's approach to the cost minimising investment programming problem. Indeed, Wright himself sees that his model is essentially a cost minimising model (Wright 1968 p.233) since he specifies both the price of the product and the maximum level of sales at those prices, so that the interdependence between price and output is effectively denied - just as it is deliberately denied for ease of computation in the Turvey model.

3. NEOCLASSICAL INVESTMENT THEORY A line of models which, it might be imagined, would cast very clear light on the investment planning problem has been associated with the neoclassical investment theorists, especially Jorgenson (see Jorgenson 1965 and 1966 and also Arrow 1964). However, these models - which are briefly outlined below produce very little of any consequence at least in the directions in which we have been following so far. Both Jorgenson and Arrow end up with what might be called "myopic" rules of behaviour - rules apply again and again for each instant of time but are otherwise independent of each other. Taking the Jorgenson model, the objective function is

maximise the present worth of a firm's profits at each instant and the constraints are:

(a) output depends at each instant on the amounts of labour and capital services provided at each instant

(b) the stock of capital services physically depreciates by the process of a constant proportion "evaporating" each instant.

The objective variable is the firm's instantaneous rate of profit, which Jorgenson defines as

$$\pi (t) = p(t)\ O(t) - w(t)\ L(t) - q(t)\ I(t)$$

the first term, $p(t)\ O(t)$, is the value of sales (price times output, with price a given constant); the second term, $w(t)\ L(t)$, is instantaneous labour cost (wage times man-hours) and the third, $q(t)\ I(t)$ is instantaneous new capital cost (price of new equipment times gross investment in capital services). The overall objective is to maximise the present worth of $\pi(t) = \int_0^T \pi(t)e^{-vt}dt$ at a given discount rate v. It clearly makes no difference however if we split up this integral into a sum of discrete instantaneous profits, because, although there are constraints on profits, none of these constraints overlap from one instant to another. Jorgenson was clearly wrong in characterising this as a dynamic model, since it is only a series of static models recurring every instant. As Turvey comments in his 1969 paper, one of the first to see this was Lucas (Lucas 1967) but the point has been forcibly made by Gould (1968) and Nerlove (1972).

It can be explained mathematically at first. The crucial variable in the analysis is the stock of capital services: $K(t)$. Jorgenson claimed he was setting up a dynamic optimization model in terms of $K(t)$, and its rate of change ($\dot{K} = dK/dt$) which is net investment. If this was the case the model would be dynamic because the optimality conditions would depend on K and $\dot{K}$ and be differential equations. But in fact $\dot{K}$ does not need to appear in Jorgenson's model because everything that could be of importance in it is summed up by the constant evaporation assumption and nothing else is effectively said about $\dot{K}$. The optimality conditions can always be reduced to ordinary static equations. The point of the critiques was that something else, besides evaporation of K, had to be assumed about $\dot{K}$, to make the model dynamic. The general solution proposed by Gould and later Nerlove was that there were positive "adjustment costs" associated with variations in the rate of investment $\dot{K}$, although neither writer is very specific on what these might be. However this is enough to bring both $K(t)$ and $\dot{K}$ into the objective function in a concrete way. The resulting optimality condition (Euler equation) is a differential equation of second order and provides a specifically dynamic model, (see mathematical appendix).

Economically the important step is to make the assumption that the rate of capacity accumulation in one period affects in a real manner the optimal rate in another period. This is exactly what Turvey's model (and the Wright model as well) does when it makes the assumptions of non-stationarity of factor costs with the resulting relationship between optimality conditions in different periods.

However before the neoclassical investment model is entirely dismissed, it is worth looking at one version of it that actually does tackle the very depreciation problem that Hotelling grappled with. This is R.E. Hall's paper "Technical Change and Capital from the Point of View of the Dual" (Hall (1968)).

Hall begins from the optimal investment decision analysis of Jorgenson and Arrow that we described above - the objective being to maximize the present worth of profits. The exercise has as an optimality condition a price equal to marginal cost condition. The Jorgenson paper developed this version of marginal capital cost as

$$vq + dq - \dot{q}$$

where v is the discount rate, d is the exponential rate of decay of capital equipment, q is the capacity cost of equipment and $\dot{q}$ is its rate of change. There is nothing 'dynamic' about this equation, nothing has to be forecast and only current variables are needed, so the cost term is not forward looking. This is to be expected from our comments above on the Jorgenson model. Hall extends this model by (a) allowing for some other form of equipment deterioration than the exponential one and (b) by

> "... relaxing the assumption of stationarity to the extent of allowing changes (in capital equipment) over time and vintage which are independent of deterioration."
>
> Hall (1968) page 37.

From our previous discussion, it can readily be understood that this substantially changes the marginal cost formula. Taking case (a) first, Hall replaces the assumption of exponential physical deterioration by using a mortality function relating the rate of decay to age of the equipment and a replacement density function which makes the probability of collapse depend on the accumulated wearing out of the equipment. The effect of this is to replace dq in the marginal cost by a term reflecting the costs associated with the probability over the remaining life of the asset of its collapse. (Hall's equation (20)). With this modification, marginal cost is made forward looking because a guess about the future mortality of the machine must be made.

A somewhat similar result obtains in case (b). Allowing for technical change (embodied or disembodied) produces a marginal cost formula in which dq is replaced by a term reflecting all the future obsolescence rates that will apply to the machine over the rest of its life. (Hall's equation (23)). Once again a guess has to be taken about the future.

The overwhelming conclusion remains: dynamic formulations of the investment decision and the marginal cost calculation have to incorporate expectations. All calculations in a dynamic model must be forward looking. This has been the important theme of all the models surveyed in this background to Turvey's programming model.

LITTLECHILD'S MODEL

Although S.C. Littlechild's model of 1970 in some senses is the most complete treatment of our problem, we need not examine it in great detail. Littlechild set out to provide a programming formulation of Turvey's 1969 model and although we shall have much to say on the success of this in the next chapter, Turvey virtually adopted the whole of Littlechild's formulation in setting out his 1971a and 1971b models which we discussed at length above. Strictly therefore the model we have been discussing could really be called the Turvey-Littlechild model. The outstanding difference is that Littlechild seeks to avoid some of the problems of indeterminateness of output which are liable to result from the iterative technique used by Turvey. Recall that Turvey - reluctant to believe planners know enough or have sufficient resources to attempt a completely interdependent model for investment planning - optimizes on the cost side, sets prices, determines demand and finally re-optimizes. We noted that this process might not converge. Littlechild's model avoids this by attempting an interdependent solution. His objective function is in terms of consumers' willingness to pay:

maximize present worth of consumers'
monetary evaluation of output less running and capacity costs

subject to the constraints:

(a) total demand satisfied cannot be less than the sum of output on all vintages of capacity

(b) output cannot exceed capacity for any vintage of plant.

(Littlechild 1970 pages 328-329, equations 13 to 15).

The comparison with the Turvey and Wright models is immediate. The results are very similar. An optimal short run pricing rule is derived with price covering running costs and amortisation charges; capacity is acquired if lifetime amortisation covers its purchase cost, and amortisation is valued in terms of a machine's quasirents which are, as expected, the dual variables associated with the constraints relating output to capacity. Littlechild brings out, as clearly as Turvey and Wright, the importance of system interdependence and non stationarity of factor prices. He incidentally provides a proof - for a myopic model of the Jorgenson-Arrow type - of the paradox pointed out by Turvey (1969, p.256) that a rise in the future rate of technical progress may raise marginal costs in the interim. This result comes out of our own analysis very easily. Recall equations (13) and (15) of chapter VI.

$$m_t = r_t^v + k_t^v \qquad \text{.......... (13)}$$

marginal cost is the sum of running cost and capital charge

$$u_t^v = \sum_{t \geq v} k_t^v = \sum_{t \geq v} (m_t - r_t^v) \qquad \text{.......... (15)}$$

the remaining stream of cost savings (capital charges) is the unamortized value of capacity. Calculate (15) for year t and year t+1 and substitute in equation (13) to get the straightforward result

$$m_t = r_t^v + (u_t^v - u_{t+1}^v)$$

where the term in brackets is the year's amortization - i.e. the fall during the year in the remaining unamortized value of capacity. Now suppose there is a rise in the rate of technical progress expected in the future; in terms of equipment which will later displace vintage v capacity, u^{v}_{t+1} the value of capacity at the end of the year, falls relatively to u^{v}_{t} so that more of this capacity's cost has to be charged immediately to this year; the result is a rise in m_t.

THE WORK OF W.A. LEWIS

W.A. Lewis is a writer whose work is being rediscovered. The discussion of optimal tariff schedules in part I of the thesis has already mentioned that not only did Lewis first discover the general solution to the peak load pricing problem (Lewis (1941)) but he was also among the first (and most lucid) to explain the theorem on pricing above marginal cost in proportion to demand elasticity when seeking to achieve a financial surplus (Lewis (1949)). It is therefore interesting to discover that Lewis was writing about optimal amortization streams and their effects on marginal cost long before the Turvey-Littlechild and Wright models appeared. His classic article on overhead costs (Lewis (1949)) first stated his point that marginal cost calculations using long lived capital equipment were based on the view that regarding output in each separate time period as a separate product implied that the appropriate model was that of joint production. This is exactly the Turvey-Littlechild point about jointness in time being equivalent to jointness in product. The definitive statement of Lewis' views on optimal amortization are contained in a neglected but important paper on depreciation and obsolescence as factors in costing (Lewis (1960)).

In this paper Lewis broadens the discussion to include both the accountant's and the economist's approach to amortization. The controversy over the appropriate depreciation policy arises from a failure to realise that depreciation is calculated for more than one purpose. He illustrates four possibilities:

(A) in order to compute tax liability,

(B) in order to evaluate prices in the second hand market,

(C) in order to determine how much of gross profit will be distributed,

(D) in order to determine the selling price of the product.

Points (A) and (C) have chiefly been the accountant's preoccupation. The crucial factor in evaluating (C) is, Lewis argues, not so much the price of the asset as the general price level and its rate of inflation. The decision maker is encouraged to maintain capital intact, and in a period of inflation will have to devote much effort to forecast general price levels and estimating depreciation on the basis of replacement cost.

It is points (B) and (D) which, being more concerned with optimal resource allocation than provident business practice (something Turvey dismisses as "good housekeeping" in his 1969 paper), are the focus of the economist's attention. It is here that Lewis returns to his classic notion of escapable cost as the basis of optimal pricing related to marginal costs. His general description is in terms of depreciation as "user costs", but his categorisation of the problem is easily recognised:

"... At some times during its life the asset will have a high yield, while at other times it will have a low yield. The firm must get what it can at different times, subject only to the condition that the asset should not be installed unless the firm expects to earn its total cost over the lifetime of the asset ..."

Lewis (1960), page 42.

There could hardly be a clearer statement of the nature of the programming model and its solution proposed a decade later by Turvey and Littlechild.

Lewis clearly realises the forward looking nature of the amortization decision. The firm must

"... make its best guess as to the projected time pattern of use of the asset".

Lewis (1960), page 43.

The yield profile is uncertain - "... it may be expected to decline progressively or to fluctuate" depending both on the evolution of costs and demand variations, as he points out.

It is not hard to argue that of all the literature which tackled the investment planning decision, Lewis' work comes closest to the modern programming approach of Turvey. Lewis is convinced of the difficulty of estimating optimal amortization in this way - his eminently quotable conclusion is

> "... the estimates are highly speculative and subjective, and are of the sort that change easily between breakfast and lunch".
>
> Lewis (1960), page 45.

It is clear from this discussion of the historical background that there has been a long history of modelling the investment decision in a way similar to that outlined by Turvey. The points which emerge again and again from the literature are

(a) the necessity of having a dynamic model - usually by making costs vary through time and with the investment decision being taken.
(b) pricing, investment and optimal depreciation are all different aspects of the same decisions: when to install and how to operate capital equipment.

INTER-RELATIONSHIPS AMONG THESE MODELS

These treatments of the global investment programming problem have certain common characteristics:

(a) they all make the attempt to construct a dynamic analysis of constrained optimization.
(b) the constraints are usually of two types:

 (i) relating output to demand or maximum output or sales
 (ii) relating output and capacity of equipment.

As far as the first characteristic is concerned, the possibility of the model being dynamic turns on the assumptions made about costs. If exogenously given factor prices do not vary over time or if they only vary in a way that means that the rate of change of capacity expansion does not systematically affect the value of the objective function, then the model remains a static one. Models which have succeeded in being dynamic have either used the embodied technological progress and rising running costs assumptions or, more generally, assume positive adjustment costs to changing investment programmes.

The interesting implications of the second characteristic relate to the Kuhn-Tucker multipliers or dual variables associated with the constraints; those corresponding to (i) yield a set of prices or - in Turvey's case - marginal costs and those corresponding to (ii) yield information on the optimum evaluation of assets and hence the ideal depreciation profile - the evaluation being in terms of profit (Wright) marginal social utility (Littlechild) or cost savings (Turvey). The following chapter considers the implications of the price/marginal cost calculations, but we can consider at this stage what the models seem to say about the shadow price evaluations of capacity.

We saw earlier that any of an infinite number of profiles might correspond to the time path of capacity shadow prices. Turvey, mainly for institutional reasons which are probably very applicable to the case of electricity supply, makes some explicit and implicit assumptions about these evaluations, (k_t^v), particularly in his 1969 paper. On new capacity, Turvey expects to see both falling capacity and running costs. On older capacity, Turvey expects running costs to rise with age and use of the capacity. Either of these assumptions leads us to expect that capacity is most useful in the early years of its life, and that the time path of the k_t^v for different years, t, of a machine's life will show a steadily declining profile. There are several possible conclusions from this. For example, it might be possible to use rule of thumb approximations to the k_t^v if we can be sure in advance of the profile that will emerge. This obviously would make decision taking a great deal easier. In his 1969 paper on 'Marginal Cost', Turvey writes that such approximations are possible; he cites a convenient approximation due to Desrousseaux: we know from our previous analysis that the crucial factor in marginal cost that reflects capacity charges is first year optimal amortization (k_t^t); Desrousseaux suggests approximating this by twice the sum corresponding to a constant annuity equivalent of total capacity cost. The implication is that capacity charges decline approximately linearly.

In preparing this thesis, I was able to discuss several of the ideas with other economists at the University of Melbourne. These aspects of approximations to k_t^v profiles were particularly interesting, and I will take the opportunity here to report some of the results that emerged from the discussions. Soper (1973) discussed the Desrousseaux approximation in particular, but found that it was not a particularly good guide, because the ratio of first year amortization to the constant annuity factor was particularly sensitive to the rate of discount and the assumed economic life of the project. (For rule of thumb approximations a guess about the most important economic variable - replacement date - has to be made.) Long lived projects and discount rates above four or five per cent destroyed the applicability of the Desrousseaux rule. Nevertheless on a very broad view, we would expect that a rule of thumb like declining balance would approximate optimal amortization better than a rule like constant sum depreciation. Where accounting conventions have to be used to approximate optimal amortization this is worth bearing in mind.

To provide some further guidance to optimal amortization, one could try to simulate versions of the Turvey model. My Melbourne colleagues, Professor L.R. Webb and Mr. B.R. Parmenter and I experimented with several simulations. The striking result - to be expected of course - is how easily the steadily declining k_t^v profile is disrupted when <u>variations in demand</u> are allowed for. The bulk of Turvey's writings on the subject are concerned with the case of steadily expanding demand which is characteristic of a mature electricity authority in a developed market; he has to rely on the variations in technical progress in capacity and running costs to provide fluctuating amortization schemes. However, in simulation terms, only a relatively small degree of demand variation can, in conjunction with technical progress in costs, produce highly variable dynamic paths for optimal amortization. (For further comments see Parmenter and Webb (1974)).

THE EMPIRICAL BACKGROUND

Turvey's model was made dynamic by explicitly incorporating the vintage capital equipment model of technological progress, with running and capacity costs falling as later vintages are installed and running costs rising as equipment ages. We know from his 1969 paper, that Turvey was a great admirer of the Salter version of the firm's decisions. While it is not strictly relevant to the theoretical development of the model, it is nevertheless worth examining for a moment the impact that the vintage model approach has had on the empirical literature.

Empirically the vintage model postulates that technological progress is "embodied" in successive vintages of equipment and plant used by the producer. Analytically this idea is very appealling but the difficulties arise when an attempt is made to measure the rate of embodied technological change in isolation from factors like disembodied technological change or equipment deterioration. Straightforward econometric models using time trends are inadequate for making the distinction.

Consequently the empirical usefulness of the vintage model has been doubted ever since the model first appeared; for the most recent survey of the criticisms and "failures" of the vintage model see Gregory and James (1973). This latter reference does acknowledge that in one industry - that of electricity generation - the vintage model may have empirical (as well as its undoubted theoretical) value. The reasons for this lie in the facts that the product is homogeneous, the technology easily identified and identification of vintage of the capital equipment is relatively straightforward.

There have indeed been several attempts to measure technical change in electricity generation; all have been carried out with U.S. data and have been related mainly to coal using plant. These are the studies of Komiya (1962), Dhrymes and Kurz (1964) and Galatin (1968). Both the Komiya and Dhrymes and Kurz studies are open to the criticism, pointed out by Galatin, that their analysis is essentially static. This arises from the relation of capital input to output in a given year. No allowance is made for an investment decision to install capital equipment to meet output requirements over a planning horizon lasting many years.

As a very brief summary of the applicability of the vintage model, we can look at the relation between the vintage of capital equipment and fuel requirements. Komiya fitted an equation of the form:

$$Y = AX^{\beta}$$

where Y is fuel input required, X is size of the generating unit and A is a fuel requirement parameter. Analysis of covariance techniques indicated that A differed significantly between vintages (though β a scale parameter did not). This was taken to indicate that newer vintages of a given scale of capacity needed significantly less fuel input.

Dhrymes and Kurz do present evidence on the impact of technological change by identifying differing technological periods. However, there is no explicit treatment of vintage models and no analysis of the degree to which technological change is embodied in latest vintages of capital equipment.

Galatin specifically tries to isolate embodied technical change using a long period investment planning model not too far divorced from the theoretical models we have considered. Embodied technical change is identified by observing shifts of the production function for machines of the same size and fuel type but of different vintage. He sums up his results as:

> "... the effects of technological change have been somewhat uneven when two adjacent vintages are compared but over the range of the sample the effect of technological change has been to save on fuel input".
>
> Galatin (1968), page 121.

This last section has been added simply to show that - at least in the electricity industry - the vintage model of embodied technological change has had an empirical as well as a theoretical impact.

CHAPTER VIII

AN INVESTMENT PROGRAMMING MODEL: THE CASE OF A GENERATING SYSTEM WITH HYDROELECTRIC AND NATURAL GAS OPTIONS, FINANCIAL CONSTRAINTS AND A TWO PARAMETER REPRESENTATION OF THE LOADCURVE

INTRODUCTION

The last two chapters were directed to setting out the Turvey model in one of its incarnations and then to investigating its relationship to the previous attempts in the literature to solve similar problems. One chief area remained to be discussed - the use of the marginal cost concepts in pricing policy. This is the subject matter of chapter IX below but in the present chapter we embark on a slightly different tack; i.e. extending the model's structure in a way that is particularly relevant to the Australian situation. At the same time an opportunity is taken to depart from Turvey's convention of allowing the time subscript to apply if necessary to hourly periods. It will be recalled that this was done to exclude the necessity of paying specific attention to the peak load problem. The model outlined below is one that would be more readily computable since the time period is assumed to be the full production year and a specific allowance is made for the peak load problem by formulating output targets in the form of a loadcurve, represented by two separate parameters.

We now attempt to build a programming model that has some especial relevance to the Australian generating situation. The development of the other electricity generating system models of course lays the groundwork and the model is entirely in the spirit and form of those already discussed. However some additional constraints are adopted to try to capture more realistically how the particular problems of Australian authorities might arise. The work of Massé and his colleagues at Electricté de France developed from models that were simple but sufficiently realistically formulated that they gave very penetrating insights.

One of the key aspects of the problem is the joint product nature of electricity output and demand. It is a load pyramid or load curve that is being produced not simply a flow of kilowatthours each identical in its economic characteristics. Massé et al. captured this by identifying three "bottlenecks" of output and optimizing with respect to those; these were winter potential, peak potential and annual output. (Massé 1962, p.167-168).

We can begin our simple model by specifying two important aspects of load generation:

A: the provision of an annual required output, measured in kilowatt-hours

B: the provision of adequate peak capacity to meet system maximum demand, measured in kilowatts.

(In the Turvey model these are specified by simply forecasting each hour's output requirements; however the two parameter representation of the load curve we adopt provides its own special results.)

In addition to the general output and capacity constraints of other models, two additional forms of constraint appear to be important in the Australian context. The first arises from the possibility of large scale use of hydroelectric schemes like the Snowy Mountain scheme for New South Wales and Victoria and the Tasmanian generating schemes. In addition South Australia and Victoria have the possibility of generating electricity using indigenous and local natural gas as primary fuel. Both these types of energy input raise the possibility of natural resource exhaustion constraints, either in the form of using up of river and reservoir sites or in the form of depletion of gasfields. Hence we must allow for finite limits on certain capacity options in the investment programme.

A further constraint, which is hinted at in other models, but is worthwhile incorporating explicitly is a financial one; there is a specific limit to yearly planned investment expenditure operating for most electricity authorities in Australia. We shall see that incorporation of this constraint amends some of the results on depreciation which other models have produced.

Finally, the model can be extended to incorporate the maintenace costs on the generating system; allowing for this, permits us to say something about the possibility of temporary or permanent retirement of plants - an aspect which only emerges as a by-product of the solution for other models.

Putting these ideas together we can set out the model schematically in the following table:

minimize the present worth of the costs of providing

A: specified annual electricity output

B: specified peak generating capacity

within the following constraints

(i) usual capacity and output constraints

(ii) finite units on the availability of using hydro and natural gas facilities

(iii) specified budget constraints.

The unknowns in the problem are of three types:

$\{Q^{sk}\}$: a sequence of installations of capacity of type s and vintage k

$\{O_t^{sk}\}$: a sequence of outputs in year t from installations Q^{sk}.

$\{R_t^{sk}\}$: a sequence of retirements in year t of some part of the installations $\{Q^{sk}\}$

The unit costs associated with these are:

c^{sk} : capacity costs of Q^{sk} per unit

r_t^{sk} : running costs of O_t^{sk} per unit

p_t^{sk} : maintenance costs of O_t^{sk} per unit

The constraints can be set up in the following manner:

(i) Output cannot be less than capacity in any year. However capacity installed is diminished throughout its life by gradual retirement. Here we are allowing for divisibility which would certainly have to be amended in a practical example:

Output is less than capacity remaining after a series of retirements:

$$O_t^{sk} \leqslant Q^{sk} - \sum_{n=k}^{n=t} R_n^{sk} \qquad \text{for all s, k, and t.}$$

here $\sum_{n=k}^{n=t} R_n^{sk}$ is the accumulated retirement which has already occurred by year t of capacity of type s installed in year k.

(ii) there must be two sets of constraints corresponding to the two parameter representation of the load curve. Firstly, the annual output requirement must be met:

Output over all types and vintages of plants must be at least A_t kilowatt-hours in year t:

$$\text{i.e.} \quad \sum_s \sum_k O_t^{sk} \geqslant A_t \qquad \text{for all t.}$$

Secondly, in each year, a specified amount of capacity must be available to meet the system maximum demand in the peak hour:

Capacity of all types and vintages in year t must be at least B_t kilowatts - allowing for accumulated retirements.

$$\text{i.e.} \quad \sum_s \sum_k Q^{sk} - \sum_{n=k}^{n=t} R_n^{sk} \geq B_t \quad \text{for all s and all } k \leq t.$$

(iii) A third set of constraints refers to the maximal rate of retirement of capacity. Accumulated retirements of parts of a given plant will always be less than or equal to capacity first installed. (In some examples this would be modified to allow for "improvements" to capacity throughout its life, but we ignore this consideration here):

$$\text{i.e.} \quad \sum_{n=k}^{n=t} R_n^{sk} \leq Q^{sk} \qquad \text{for all s and all k, and t.}$$

(iv) The first of our completely new constraints relates to exhaustion of facilities. Retirements can be ignored here because they cannot regenerate depleted gas reserves or recreate natural river or reservoir facilities. Hence we have only a simple relationship between hydro and natural gas capacity and the finite limits to it.

Installed hydro capacity (or natural gas using capacity) has a given upper limit:

i.e. $\sum_k Q^{hk} \leqslant L_h$: superscript h refers to hydro facilities.

$\sum_k Q^{gk} \leqslant L_g$: superscript g refers to natural gas facilities.

An alternative way of representing these limits - more applicable to Q^{gh} than Q^{hk} - would be to allow the authority to buy flows of natural gas or to rent water facilities at steadily rising charges, as finite limits are approached.

(v) The other completely new set of constraints limits yearly investments expenditure to a fixed annual appropriation. (Of course several other ways of formulating financial constraints could equally apply):

i.e. total expenditure on latest vintage plant of all types is less than or equal to a financial limit in each year:

$\sum_s c^{st}Q^{st} \leqslant D^t$ for all t.

Where D^t is year t's budget for capital expenditure. It is interesting that in a sense D^t sets the most rigid boundary on any solution. Massé (1962) reports that for low values of D^t in postwar France optimal investment programmes were limited to use of conventional thermal plant: however as soon as higher values of D^t

allowed for some hydroelectric options, the cost savings were so large that the optimal solution quickly converged on a heavy bias towards hydroelectric plant.

(vi) Finally of course negative values of our unknowns are not permitted.

The solution of the model proceeds straightforwardly now, and we can concentrate on examining the effect of the new constraint formulations on the results already established in the literature. To begin with, the Lagrangian function of objective and constraints is set out below. In doing so we have introduced a series of Kuhn Tucker multipliers:

$$(\gamma_t^{sk}, \quad \alpha_t, \quad \beta_t, \quad \sigma_t^{sk}, \quad \phi, \quad \psi, \quad \delta_t).$$

$$\text{Minimize } C\left[\{Q^{sk}\}, \{O_t^{sk}\}, \{R_t^{sk}\}, \{\gamma_t^{sk}\}, \{\alpha_t\}, \{\beta_t\}, \{\sigma_t^{sk}\}, \phi, \psi, \{\delta_t\}\right]$$

$$=$$

$$\sum_s \sum_k \left[c^{sk}Q^{sk} + \sum_t \left(r_t^{sk}O_t^{sk} - p_t^{sk}R_t^{sk}\right)\right]$$

$$+ \sum_s \sum_k \sum_t \gamma_t^{sk} \left[O_t^{sk} - Q^{sk} - \sum_{n=k}^{n=t} R_n^{sk}\right]$$

$$+ \sum_t \alpha_t \left[A_t - \sum_s \sum_k O_t^{sk}\right]$$

$$+ \sum_t \beta_t \left[B_t - \sum_s \sum_k \left[Q^{sk} - \sum_{n=k}^{n=t} R_n^{sk} \right] \right]$$

$$+ \sum_s \sum_k \sum_t \sigma_t^{sk} \left[\sum_{n=k}^{n=t} R_n^{sk} - Q^{sk} \right]$$

$$+ \phi \left[\sum_k Q^{nk} - L_n \right]$$

$$+ \psi \left[\sum_k Q^{gk} - L_g \right]$$

$$+ \sum_t \delta_t \left[\sum_k c^{st} Q^{st} - D^t \right]$$

In analysing the solution using the theorems of non linear programming, we will take as read the usual "complementary slackness" results whereby either the necessary conditions hold with equality or the corresponding variable takes a value of zero. In passing it can be pointed out that we have not included in the problem the specification of any inherited capacity since that is not an aspect on which we are going to focus attention.

We can begin by examining the partial derivatives of the optimal value of C with respect to capacity and output. Taking capacity first, we write, for all vintages and types actually installed (temporarily leaving aside hydro and natural gas facilities).

$$\frac{\partial C^*}{\partial Q^{sk}} = 0 = c^{sk} - \sum_t \gamma^{sk} - \beta_k - \sum_t \sigma_t^{sk} + \cdot \delta_k \quad \ldots\ldots\ldots (1)$$

$$\frac{\partial C^*}{\partial O_t^{sk}} = 0 = \gamma_t^{sk} + \gamma_t^{sk} - \alpha_t \quad \ldots\ldots\ldots (2)$$

These two equations correspond to the pricing or costing and investment rules already discussed in the literature. Taking equation (2) first of all we have:

$$\alpha_t = r_t^{sk} + \gamma_t^{sk}$$

and applying our usual analysis, α_t can be interpreted as:

$$\frac{\Delta C^*}{\Delta A_t}$$

which is the effect on the optimal value of costs of a small increment in annual power requirements: i.e. it has a cost per KWhour dimension and of course reflects, as we know, a running charge and an amortization charge calculated over a mix of old and new equipment of differing types; γ_t^{sk} reflects the amortization charge or cost savings due to the presence of a small extra amount of capacity of the appropriate type.

Turning now to equation (1) we see how these amortization charges recover the costs due to equipment installations. However the expression here is much more complex than the usual case.

$$\Sigma \gamma_t^{sk} + \Sigma \sigma_t^{sk} + \beta_k = c^{sk} + \delta_k$$

The term on the right is the sum of the price of an item of new equipment <u>and</u> the premium which must be added to it reflecting the budget constraint D^k operating in that year. The market price of an item of new capacity does not reflect its true cost to the authority because the authority has a resource shortage (in capital funds) which the equipment market does not recognise. It is this resource cost ($c^{sk} + \delta_k$) which must be recovered by amortizing capacity. Only if capacity is worth enough to recover this augmented resource cost will the authority invest in it. The presence of the constraint on capital expenditure adds another dimension to the problem of "time" depreciation. It is no longer the case that amortization in the sense used by Hotelling, Turvey and Wright(1) - of a fall in the value of cost savings of equipment - is a sufficient measure of ideal depreciation. The authority now has a capital expenditure constraint which compels it to consider problems of how to maintain its capital intact in the traditional ways discussed by economists - notably by recognising the difference between the historical and replacement cost targets for a depreciation charge. Lewis (1960) provides a good discussion of how different objectives in depreciation policy lead to different concepts of the theoretically ideal depreciation method. On the left hand side we have a measure of the benefits or cost savings attributable over its lifetime to a unit of capacity of the type and vintage in question. We can recognise the first term immediately: it is the cost savings that capacity provides in producing for annual power requirements (A_t): i.e. the life-time sum:

---

(1) see the discussion in Chapter VII above.

$$\Sigma_t \gamma_t^{sk}$$

but this capacity has other benefits as well. The term $\Sigma_t \sigma_t^{sk}$ is again a present worth of lifetime cost savings this time associated with early retirement of capacity; as we show below retiring capacity early may produce cost savings which will help to amortize the resource cost of that capacity. The third term on the left hand side of the equation is:

$$\beta_k$$

and from looking at the constraints we know:

$$\beta_k = \frac{\Delta C^*}{\Delta B_k}$$

the cost variation associated with a change in demand at system peak - i.e. in the hour of system maximum demand. Hence the cost savings associated with capacity are of three types:

(i) $\Sigma_t \gamma_t^{sk}$: cost savings for annual power requirements

(ii) $\Sigma_t \sigma_t^{sk}$: cost savings from varying retirements rates

(iii) β_k : cost savings from providing extra peak capacity.

These will vary with the nature of equipment types and the importance of this model is the allowance it makes for hydroelectric schemes in the plant mix. The reason for this is that hydroelectric schemes provide one of the few options for "storing" power output either in seasonal or in daily reservoirs. Wherever capacity can be built up -

by winter rainfall or by night-time pumping up of water levels - for a specific period, then that capacity comes into its own in meeting peak demands over and above base load operation. Hydroelectric schemes have a great deal more of this sort of peak flexibility than - say - conventional thermal plants which usually run for base load at almost full capacity. We have therefore the important result that in evaluating marginal increments to capacity in the form of a mix of for example

conventional thermal capacity

nuclear capacity

or reservoirs,

it is not enough to look at costs only on a kWhour basis because the third option (reservoirs) - where they are feasible as in Australian electricity supply - are likely to have a large cost saving item on a kW basis at times of system maximum demand or large seasonal demand. It is this form of cost saving that is measured by β_k.

We can now take up the comments we made about cost savings due to retirements. Looking at the objective function we can see that the retirements enter total system costs with a negative sign because they save on maintenance expenses

$p_t^{sk} R_t^{sk}$ = cost saving in form of reduced maintenance on a fraction of retired capacity.

For capacity retirements that do occur the appropriate partial derivative is:

$$\frac{\partial C^*}{\partial R^{sk}} = 0 = -p_t^{sk} + \sigma_t^{sk} + \beta_t$$

and we can examine the cost saving from early retirement - the term σ_t^{sk} reflecting a relaxation in the retirement constraint - as follows:

$$\sigma_t^{sk} = p_t^{sk} - \gamma_t^{sk} - \beta_t$$

Quite simply it does pay to temporarily or permanently retire a unit of capacity if the saving on maintenance expenses (p_t^{sk}) exceeds the cost savings in power output and peak capacity provision (γ_t^{sk} and β_t) which that capacity provides. This condition really only determines temporary retirements; nevertheless reasonable assumptions about the time profile of p_t^{sk}, β_t and γ_t^{sk} might suggest that - unless demand is extreme in its fluctuations - a machine once retired will stay that way. This formulation has the advantage over other models of making explicit the decision to retire a part of capacity.

So far in our discussion we have not made use of the constraints on exhaustion of resources. If, for example, we focus attention on investment in further natural gas facilities our investment criterion is the special case of equation (1) for k = g:

$$c^{gk} - \sum_t \gamma_t^{gk} - \beta_k - \sum_t \sigma_t^{gk} + \psi + \delta_k = 0$$

and solving for ψ we obtain

$$\psi = \sum_t \gamma_t^{gk} + \beta_k + \sum_t \sigma_t^{gk} - c^{gk} + \delta_k \qquad ..... (3)$$

i.e. the surplus of benefits over costs for investment in natural gas facilities. By our usual reasoning ψ itself represents the effect on optimal system costs of a small change in the overall availability of natural gas facilities

$$\psi = \frac{\Delta C^*}{\Delta L_g} \qquad \ldots\ldots (4)$$

Thus from equation (4) we know that ψ measures - in terms of cost savings - the benefit of additional natural gas discoveries, and, hence, is a measure of the attractiveness of exploratory effort in natural gas. This benefit is in turn (equation (3)) the difference between the present worth of a stream of cost savings from having extra natural gas facilities and the resource costs of installing these facilities. We can make use of equation (3) to throw some light on the problem of natural gas reserves. It is often superficially stated that a low rate of additions to reserves heralds depletion of possible resources. In fact the process goes as follows. Low demand for the final product is expressed in equation (3) by low values for $\Sigma_t \gamma_t^{gk}$ and β_k, the value of additional facilities; this translates into a low figure for the incentive to explore: ψ, and hence a low figure for additions to reserves. In fact the true world situation is not nearly as integrated as this reasoning suggests but it at least pictures the effect that demand variation has on every stage of the production process.

The same reasoning of course applies to the case of hydro facilities. Using equation (3) in this case we have as the investment rule:

$$(c^{hk} + \delta_k + \phi) = (\sum_t \gamma_t^{hk} + \beta_k + \Sigma\, \sigma_t^{hk})$$

Which states that for the special category of hydroelectric investment, resource costs of capacity - which must be covered by the positive benefits of the investment - are of three kinds:

c^{hk} : capacity installation cost

δ : the shadow cost of the restricted funds in the budget

ϕ : the shadow cost reflecting the finite limit on reservoir sites which grow more and more scarce.

This last item ϕ , which can be interpreted:

$$\phi = \frac{\Delta C^*}{\Delta L_h}$$

will eventually tell against hydroelectric schemes and motivate a re-direction towards other types of capacity to fulfil base and peak load requirements. The values taken on by ϕ will therefore be one of the important factors determining the dates on which Australian electricity authorities begin seriously to consider the adoption of nuclear generating capacity.

Quite clearly our model could be further developed and there remains the large area of transmission possibilities and constraints to be included in an actual application. Nevertheless the model we have presented here shows fairly well some of the considerations essential in global planning of Australian electricity supply; one of the most helpful products of the cost minimizing programming exercise is the insight it can give into problems like the incentive to look for new natural resource stocks or the timing of the introduction of radically new capacity options.

CHAPTER IX

THE IDEA OF MARGINAL COST IN INVESTMENT PLANNING MODELS

INTRODUCTION

From our initial discussion of investment planning models in general and Turvey's model in particular we know that one of the most important objectives in planning is to provide an estimate of the cost bases on which prices may be constructed. Indeed, we saw that Turvey regarded this information on costs as the main achievement of the modelling process. The purpose of this chapter is to consider what we can learn about marginal costs using investment planning models. Two points are of outstanding initial importance. Firstly, basing the measurement of marginal cost on an optimization exercise means that the results will be forward looking and related to forcecasts and expectations rather than backward looking historical measurements. Secondly, we shall have a variety of marginal cost measures to choose from. At the end of this chapter we will have drawn together the results of several models of the pricing process. It is these results that formed the basis of the theoretical introduction to Part I of this thesis (on empirical estimation of tariff policies and costs). One of the measures of marginal cost that we examined in Turvey's model of Chapter VI was

$$\frac{\Delta C^*}{\Delta X_t} = m_t$$

i.e. the effect on optimal system costs of a small change in forecast demand in year t. However even here we cannot say that we have a unique measure of marginal cost. There is a different value for m_t for every discrete time period in the planning model's horizon. Which of these corresponds to marginal cost? The complexity of the concept "marginal cost" goes much deeper. We have looked only at marginal cost of output increments - but we could just as easily look at marginal cost of financial constraint increments, resource exhaustion changes and a host of other small variations in the authority's environment. If we try to relate the programming measure, m_t, to textbook definitions of marginal cost how are we to treat the well established distinction between short run and long run marginal costs and what can we say about the relationship between these at the otpimum? Given these considerations, do we have an adequate measure of marginal cost for pricing policy? All these are questions that Turvey set out to tackle in developing an optimization approach to marginal cost measurement.

THE DUAL VARIABLE MEASURE OF MARGINAL COST

We can begin the discussion by looking at the marginal cost concept developed by Turvey in 1971

$$m_t = \frac{\Delta C^*}{\Delta X_t}$$

Turvey has written in some detail about the meaning of this equation (Turvey 1971a, pages 53-58). Suppose we have set up a cost minimizing model like that of Chapter VI or Chapter VIII in terms of a constrained optimization programme; there are several outstanding characteristics of the formulation of which we ought to be aware. Of course we have the exogenous variables of the model

- the forecast demand requirements $\{X_t\}$ from the present stretching perhaps to infinity.
- the inherited capacity of the electricity authority or investment planning authority.
- a whole battery of expectations about future running and capacity costs and the effect on them of technological progress - emphasising the important non stationary character of the model.

Given all these we have a series of recurring decisions to take, planning the further acquisition and utilization of capital equipment in each year of the programme ahead; the most important of all the decisions are those relating to the present year's acquisitions and operations, because - while all the other years' decisions remain hypothetical - the present year's decisions are about to be put into practice. Thus we have a picture of a situation in which a stream of forecasts will cause the authority to start off from its inherited position on a new sequence of decisions and operations. Suppose now that within this long time scale of forecasts the demand requirement for the product in one future time period changes by a very small amount from its previously forecast level to a new forecast level.

demand in year s changes from a forecast of X_s

to a forecast of $(1 + \Delta)\ X_s$.

The least expensive way of meeting this demand increment or fall - ΔX_s - will cause a widening ripple through the investment and operating plans for every year ahead of the whole programme: the present worth of total system costs changes by a very small amount as a result of a whole sequence of plan changes:

e.g. some capacity planned for construction in year s-2 may need to be constructed in s-3 while some scheduled retirements for s+4 may be amended.

Because of the importance of system interdependence and the dynamic effects of technological progress the effect on total system costs - C - of ΔX_s may be very different from the simple addition of the costs of running capacity a little harder in year s; the latter option could of course be chosen but may not be the cheapest option available and we would expect a systematic analysis of the cost of the marginal change to relate to the alteration in costs at the optimum. This is the rationale behind the careful examination of the dual variables (m_t, k_t^v, etc.) in the model - so that the optimal effects on total systems costs of ΔX_s may be calculated. This is so because we know that under general conditions the dual variables reflect small variations at the optimum in the exogenous variables entering the constraints on the problem. Clearly there arises here the multiplicity of marginal costs since there are so many exogenously given variables in the model; any potential variation in any one of them would produce an alteration in total system costs at the optimum.

Of course for tariff purposes we are primarily interested in the cheapest way of meeting small variations in demand requirements, hence it is the dual variables associated with $\{X_t\}$ - the m_t - which attract our attention as the basis of tariff schemes. The first occasion on which Turvey emphatically drew attention to the necessity of calculating a forward looking measure of marginal cost for tariff setting purposes was in his 1969 paper. The programming model which later developed was initiated by an article by Littlechild (Littlechild 1970). Littlechild attempted to give a programming formulation to the discursive approach that Turvey had set out. Hence the dual variable measures of marginal cost - m_t - are really to be seen as programming translations of Turvey's pathbreaking and penetrating ideas contained in that 1969 paper. From our discussion above we know how to interpret these dual variables - however the question remains open whether these measures adequately reflect the marginal cost concepts that Turvey had originally set out.

TURVEY'S HISTORICAL DYNAMIC MARGINAL COST

The ideas on marginal cost associated with Turvey's work were first sketched in his book on electricity supply in 1968 and later given a larger development in his 1969 paper. Turvey was at pains to relate marginal costs to the optimal planning of capacity:

> "... what is important however is that the estimation of long run marginal cost is consistent with the planning of capital expenditure. Both require the same calculations ... "
>
> (Turvey 1968, page 51)

He wanted to emphasise the long term planning aspects of the twin problems of capacity accumulation to meet load requirements and the pricing of those requirements. In a section of the 1968 book on "the time pattern of marginal cost" (pages 54-55) he discusses the consequences of the emergence of a permanent load increment. A new demand requirement appears and it is known that this is going to continue for the most if not all of the remaining planning horizon. One response of the authority might be to find the constant annuity equivalent of the costs of meeting this permanent load increment and to set up a long term contract for charging for it. He dismisses this idea however, firstly because electricity is rarely sold in long term contracts and secondly, because of the arbitrariness of using annuity equivalents. A better response he suggests is to set up an alternative definition of marginal cost. We attempt to "isolate the share of the increment in system costs caused by a given permanent load increment which can be attributed to a particular year, n," as:

" ... (i) the present worth of the increment of system costs resulting from a permanent load increment starting at the beginning of year n

less

(ii) the present worth of the increment of system costs resulting from the same permanent load increment starting at the beginning of year n+1 ..."

(Turvey 1968, page 55)

This is a very interesting definition of marginal cost and a great deal ahead of the typical textbook definition of the first derivative of a continuous cost function related to output: $C'(X)$. In spelling out the ingredients of a measure of marginal cost defined in this way, Turvey discovers that at least three aspects need careful calculation:

(i) the present worth of fuel savings associated with new capacity, if any

(ii) the distinction between costs associated with annual power increments and those associated with additions to peak capacity

(iii) the exact nature of the capital charge involved.

Now clearly the programming models discussed - say in Chapter VIII- do allow for these three aspects; nevertheless it is not apparent that the definition of $\Delta C^*/\Delta X_t$ covers all that Turvey appears to be looking for in his definition above. Summing up: Turvey gives a complex definition of marginal cost as a tariff basis and indicates some of its ingredients; the programming method's dual variable also captures some of the ingredients but is not a definition couched in terms of the potential starting dates of permanent load increments. It seems just possible that Turvey's interesting 1968 definition has not been completely analysed.

In 1969 in his paper on "Marginal Cost" Turvey develops his radically new definition verbally and diagrammatically. His starting point is to

> "... survey recent developments in the concept of marginal cost with particular reference to marginal cost pricing ..."
>
> (Turvey 1971a, page 282)

He wants to emphasise that the traditional textbook treatment is not at all useful for practical applications - it is essential to develop a historical dynamic concept - a measure which allows for problems like system interdependence and amortization calculations. More significantly this concept needs to be usuable enough to allow economists to

> "... act as would be insiders who are trying to decide what ought to be done ... "
>
> (Turvey 1969, page 282)

He does however point up one textbook result that is worth examining in any model: "marginal short and long run costs coincide when capacity is optimal", (Turvey 1969, p.293). This theorem encapsulates so much complex analysis that it would clearly be a very useful result to carry about in practical calculations. (Turvey himself, in a much earlier paper, (Turvey 1964) provides an extension to the case of tariff setting and capacity installation in a very simple model of the electricity supply situation.)

Consequently his approach to the problem of surveying recent work on marginal cost is twofold: can we broaden the textbook treatments to give practical answers to tariff setting problems - and can we still condense our results into useful theorems about short and long run marginal costs?

He begins - in his 1969 paper - his attempt on the first part of the problem by reviewing the contributions of other writers who have been worried by the lack of regard paid in the textbook treatment to problems like the fixity and longevity of capital equipment and the diffusion of technological progress. In particular he admires the work of Salter, which we have already reviewed in Chapter VII above. While these other contributions would make a worthwhile study in themselves, we can quickly summarize the main results as emphasising that

> " ... cost and output have time dimensions ... "

furthermore cost and output will depend on the extent to which capacity is inherited from the past; all this says that

> " ... a cost analysis which is to be useful in decision making needs to be historical dynamics, not comparative statics ... "
>
> (Turvey 1969, page 287)

Turvey then proceeds to develop a definition which he thinks fits the requirements - and it is completely in sympathy with the measure he had earlier developed for the case of electricity supply and which we examined a few paragraphs above.

Two features of the definition are important:

(i) the use of the definition is in "... making decisions about pricing ..."

(ii) "... the nature of the concept required can thus be ascertained only with reference to the objective function, to the constraints and to the amount of information available ..."

Turvey then makes a point earlier made by Boiteux (1951): many price or tariff structures - particularly for a public enterprise's products - cannot be changed frequently. In conjunction with the fact that consumers' decisions are frequently long term ones, this means that

> " ... the marginal cost concept which is relevant to pricing decisions consequently relates to permanent output changes ..."
>
> (Turvey 1969, page 288)

Turvey then gives his definition of marginal cost under these considerations.

The definition is given in the context of an example of an enterprise planning for a PERMANENT OUTPUT INCREMENT - an increment which is postulated as being "large enough to be noticeable but small enough to be marginal". The permanent output increment could begin in any one of the future years of the planning horizon, and marginal cost is based on the differences made to the present worth of system costs by different starting dates. A diagrammatic illustration is given in Figures IX,1 and IX,2. We can tabulate the crucial variables as follows i.e. the values for the present worth of system costs and the output stream being costed. In particular consider the following calculations.

TABLE IX, 1

NO.	OUTPUT STREAM							COST VALUE
(1)	X_0	X_1	...	X_{S-1}	X_S	X_{S+1}	X_{S+2}	PWC(I)
(2)	X_0	X_1	...	X_{S-1}	$X_S+\Delta$	X_{S+1}	X_{S+2}	PWC(II)
(3)	X_0	X_1	...	X_{S-1}	$X_S+\Delta$	$X_{S+1}+\Delta$	$X_{S+2}+\Delta$	PWC(III)
(4)	X_0	X_1	...	X_{S-1}	X_S	$X_{S+1}+\Delta$	$X_{S+2}+\Delta$	PWC(IV)

Output stream (1) can be regarded as benchmark: it shows a forecast level of output requirements proceeding from year zero onwards:

$$X_0 \quad X_1 \quad X_2 \quad \ldots \quad X_{S-1} \quad X_S \quad X_{S+1} \quad X_{S+2} \quad \ldots$$

Output stream (2) shows the same forecast output levels for all years except year S; in year S forecast output is Δ units higher

$$X_S+\Delta$$

and output then returns to its previously forecast level.

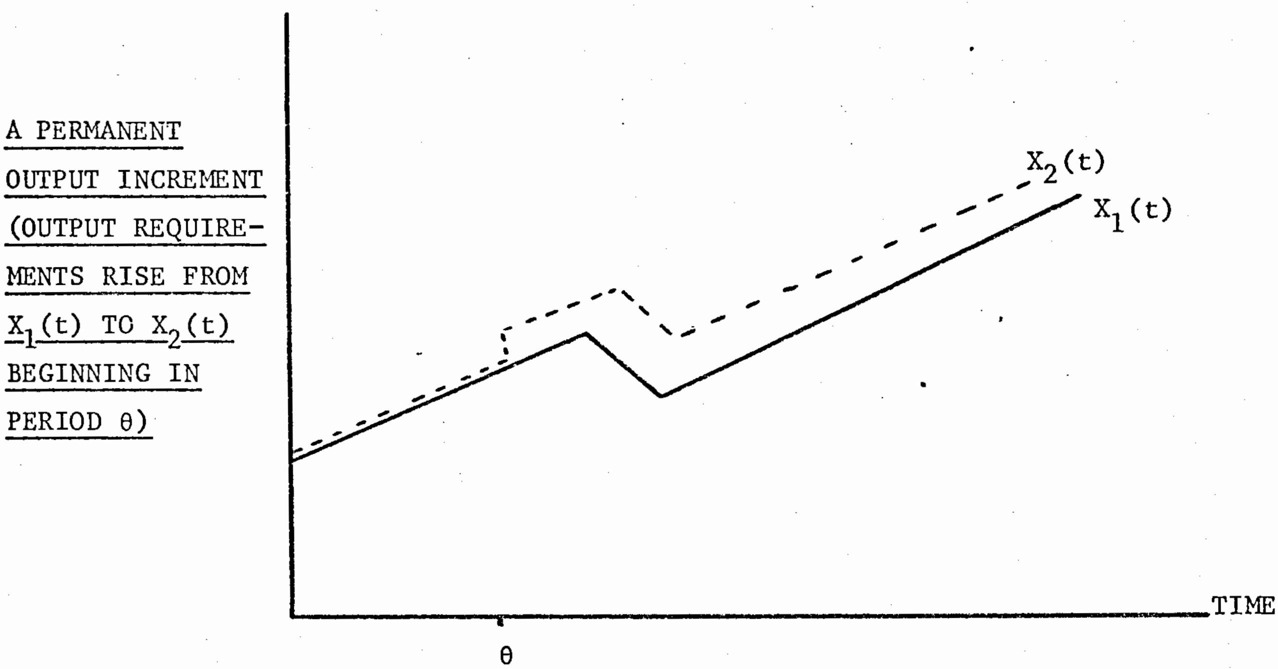


Figure IX,1

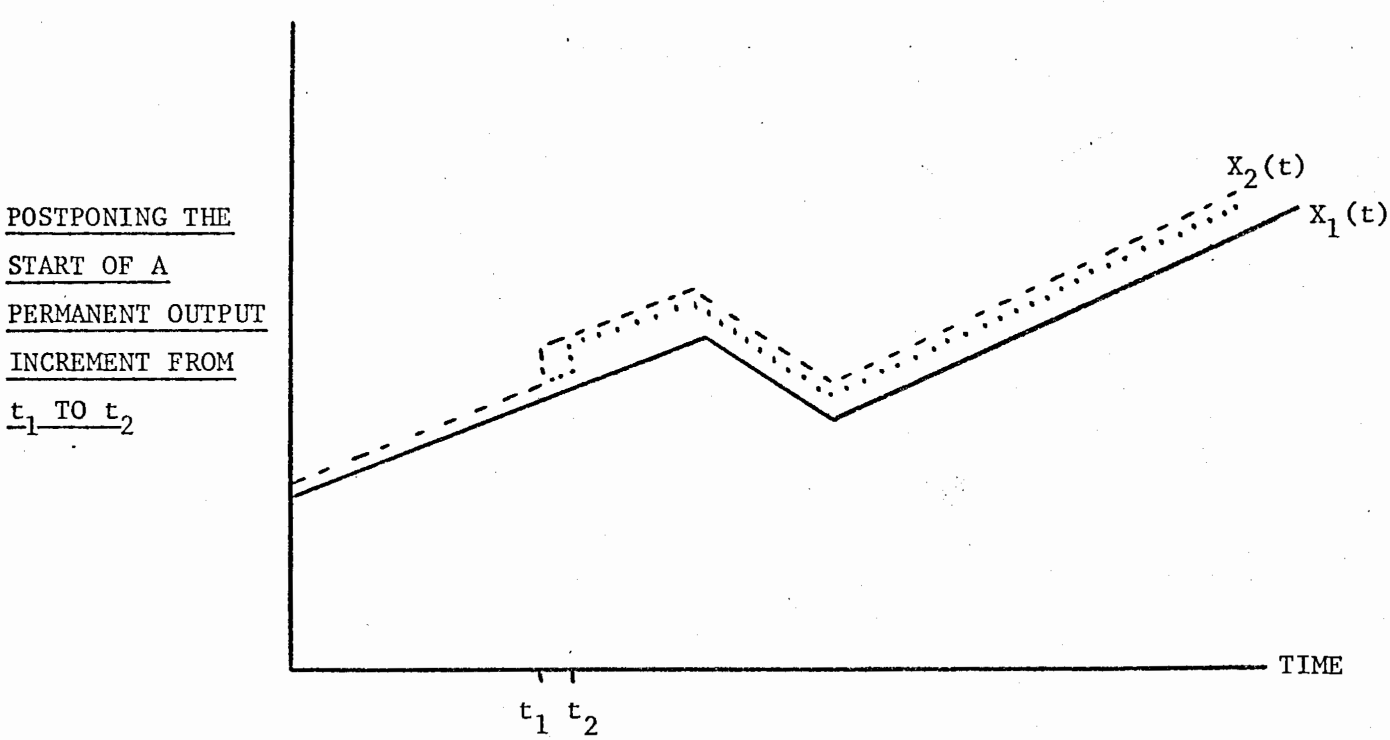


Figure IX,2

If we compare the associated system cost values we have a definition of marginal cost corresponding to one earlier Kuhn-Tucker analysis:

$$\text{PWC(II)} - \text{PWC(I)} = m_S = \frac{\Delta C^*}{\Delta X_S}$$

However, although this is the marginal cost concept that figures so importantly in the Littlechild and later Turvey models, it is not the basic definition of Turvey's 1968 book or 1969 paper.

Consider output stream (3); this shows a permanent output increment beginning in year S and carrying on

$$X_0 \quad X_1 \quad X_2 \quad \cdots \quad X_{S-1} \quad X_S{+}\Delta \quad X_{S+1}{+}\Delta \quad X_{S+2}{+}\Delta \quad \cdots$$

Suppose this permanent output increment began one year later; the output stream would be (4)

$$X_0 \quad X_1 \quad X_2 \quad \cdots \quad X_{S-1} \quad X_S \quad X_{S+1}{+}\Delta \quad X_{S+2}{+}\Delta \quad \cdots$$

with cost level PWC(IV). The marginal cost concept Turvey is interested in can then be written simply as

$$\text{PWC(III)} - \text{PWC(IV)}$$

> " ... marginal cost for any year is the excess of (a) the present worth in that year of system costs with a unit permanent output increment starting then, over (b) the present worth in that year of system costs with the unit permanent output increment postponed to the following year ... "
>
> (Turvey 1969, page 289)

and this measure is clearly going to coincide with m_S in only a restricted number of cases. This 1969 definition mirrors exactly the measure used by Turvey in his essay on the electricity industries problems which we examined above, but it is clearly of a different nature to the dual variable measure we have been examining. If we take these two threads of marginal cost calculation we can follow them out to some interesting insights into electricity planning, but we have to begin by clearly distinguishing the two results.

Turvey himself is not quite explicit enough on the relationships between the two measures - in his 1969 paper he simply summarises:

> " ... while the concepts presented here are general ones, actual measurement requires a cost model specific to the industry concerned; (...) a programming analysis may be used which has the great advantage of directly producing marginal cost estimates as the duals of the output constraints ... "

(Turvey 1969, page 290)

In the following pages we try to work out some of these cost concepts and we can indicate our main results in the following way:

(i) Turvey is trying to set a price for electricity and this leads to his 1968-1969 definitions.

(ii) the global programming model provides (in m_S) a measure of what looks like the accepted idea of marginal cost: $\Delta C^*/\Delta X_S$ and at the same time incorporates all those crucially important factors of technological progress, dynamic amortization and expectations which Turvey wants to emphasise is the basis of electricity planning.

(iii) the most important thing to understand is that we have left behind the textbook world of static long run and short run scale cost curves.

Let us begin with point (iii), it is not hard to get the impression from Turvey's 1969 paper that he has tried to maintain the concepts of long run and short run marginal cost in his analysis. The object of the analysis of investment planning is to achieve the optimal adjustment of capacity to demand and this usually is associated with an equality between short and long run marginal costs. However having said this, Turvey appears to leave some doubt about the nature of his definition. His concept appears to be a long run idea because of the optimal capacity notion and the permanent nature of the output increment; nevertheless, he says, it is misleading to consider his definition as a long run concept.

Kay took up Turvey's ideas in his 1971 comments and criticised the analysis relating to the optimal capacity theorem. Turvey, in his reply, appeared to accept Kay's strictures by simply admitting

"... my paper was also a little confused about LRMC versus SRMC pricing ... "

(Turvey 1971b, page 371)

There are two points at issue in the Kay-Turvey exchange: the optimal capacity theorem and whether Turvey's concept of marginal cost can be said to be a long run measure.

The derivation of the optimal capacity theorem is set out in any textbook. It is summed up in a well-known diagram which is reproduced below (figure IX,3). The figure shows a short run marginal cost curve (SMC) and a long run marginal cost curve (LMC). Short run marginal cost depends not only on factor prices and output but also the fixed availability of one or more inputs (here $\bar{M}$ machine hours). The long run marginal cost curve is constructed without any such input fixity. If output was to be at the level $\bar{X}_A$ then the fixed short run availability of machine hours is exactly that which in the long run minimizes costs of production. This is not the case for output $\bar{X}_B$. For small variations in output around $\bar{X}_A$ it makes no difference whether existing capacity is worked harder (or less hard) or new capacity is installed, since the costs of either action are identical. Consequently we can regard capacity as being at the optimal level for output $\bar{X}_A$.

We have therefore the general result that when capacity is optimally adjusted to demand (or required output level) short run marginal cost equals long run marginal cost.

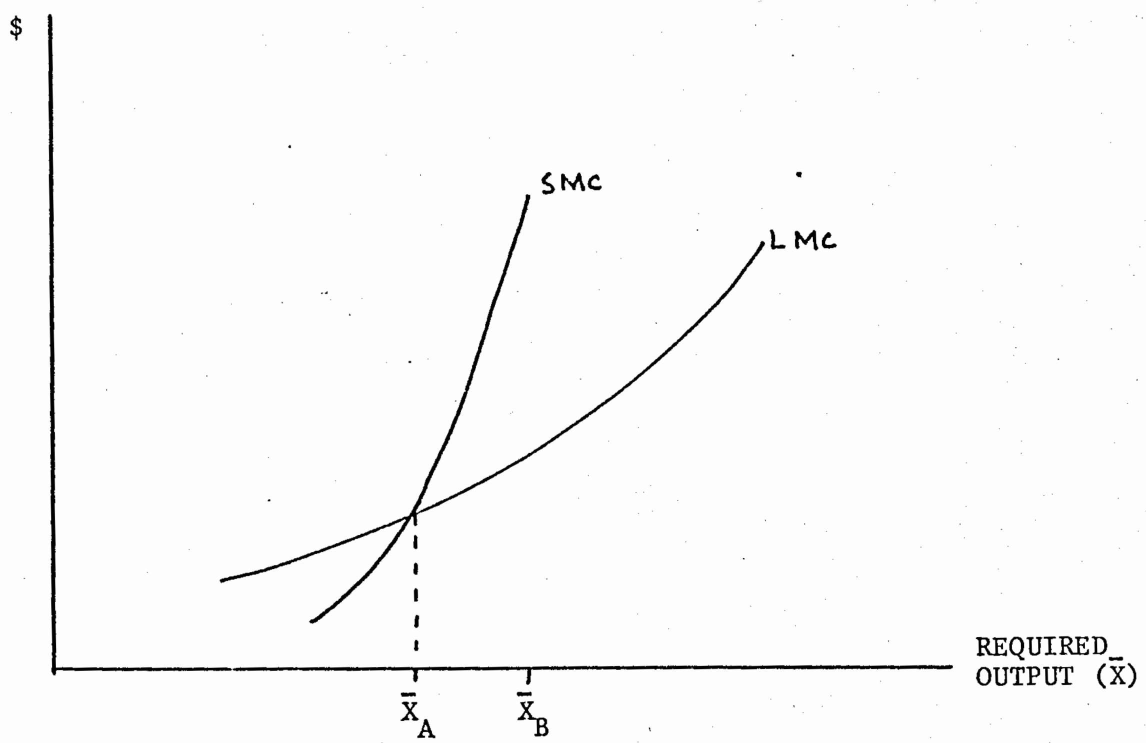


Figure IX, 3

$$SMC = F(w, r, \bar{X}, \bar{M})$$

$$LMC = G(w, r, \bar{X})$$

r, w = factor prices; $\bar{X}$ = output requirements

$\bar{M}$ = fixed amount of machine hours available for the short run case illustrated

It is not hard to see how this carries over to our programming model. We have already noted that this gives both a costing and an investment rule. Referring back to figure VI,4 of Chapter VI we see how different types of machine are incorporated into the production flow for each period. At the optimum we can compare the newest and oldest machines in use; one is on the verge of being scrapped with low quasi-rent, while the other is newly in use with large quasi-rent, but an increment to output requirements could be met from either machine at the margin; this is simply a reflection of the textbook choice of working old capacity harder or installing new capacity.

Turvey sees that this will be the case in his investment planning model but in the context of his reframed definition of marginal cost in terms of permanent output increments is reluctant to state the proposition in long run-short run cost terms. The optimal capacity result is "no more than a truism" in the context of an investment planning model.

It is Kay who tries to translate Turvey's permanent increment cost concepts into the textbook long run-short run distinction.

He sets out the four output streams that we used above in table IX, 1, and he asserts the following defintions:

(a) short run marginal cost is PWC(II) - PWC(I)

(b) long run marginal cost as defined by Turvey is PWC(III) - PWC(IV) assertion (b) is valid - it is the description of Turvey's concept we illustrated above. Using hypothetical cases for situations I to IV Kay then dismisses the optimal capacity proposition - the equality of short run and long run marginal costs is neither necessary nor sufficient for it. However we can see immediately what Kay's error is; no one - least of all Turvey - has claimed that

PWC(II) - PWC(I)

is short run marginal cost. Turvey perhaps causes some confusion by inserting into his definition of permanent output cost concepts the discussion of optimal capacity in the traditional textbook terms but he would have been the first to recognize that short and long run distinctions do <u>NOT</u> relate to the time length or magnitude of the output increment. The short run is distinguished from the long run by the fixity of one or more factors. Kay's counter-example is based entirely on a misconception of what short run marginal cost is.

Now we showed above how the optimal capacity theorem of the textbooks can be coaxed out of the investment programming model. What can we say about the short-long run distinction in the context of Turvey's permanent output increment ideas?

About the best thing we can do is to jettison the whole long run-short run distinction. We can make it clear from the start that we are dealing with optimally adjusted investment programming models incorporating various constraints and operating through time. This is a complete description; to bring in the textbook notions of long and short run is a misleading red herring.

Having established this, we want to investigate how Turvey's permanent output increment concept is to be evaluated in an investment planning model. We achieved a measure of marginal cost in the form of the dual variable m_S but have already decided that it perhaps does not contain the full richness of the 1969 Turvey definition of marginal cost in terms of the postponement of the starting date of the permanent output increment. What we require is a reformulated investment planning model that will distinguish between permanent and temporary output changes.

PERMANENT AND TEMPORARY OUTPUT CHANGES IN A DYNAMIC INVESTMENT PLANNING MODEL

In this section we are going to tackle head on Turvey's 1969 definition of marginal cost. We want to construct a dynamic planning model for investment decisions that will allow us to examine permanent output increments and the costs that these impose and then to examine the cost changes at the margin of varying the starting dates of these output increments.

French electricity enconomists have been particularly active in developing such models and much of our analysis will be based on a recent though rather difficult French model. As far as we are concerned the crucial results will be in answer to the questions:

(a) does Turvey's 1969 definition of marginal cost translate into expressions for running and capacity charges with which we are familiar?

(b) what implications for pricing policy does the model throw up?

In 1970 two economists of Electricité de France - M. Albouy and J.C. Nachtigal - set up a continuous time model of electricity investment planning which contains enough to tackle Turvey's marginal cost definitions. Turvey himself refers to the article in his 1971 book but does not use its interesting results. The French model is a very rich piece of work but part of its richness has to be diluted if we try to model Australian or British electricity policy in it. Both in practice and in analysis Electricité de France has considered one of its policy options to be the imposition of selected supply interruptions on consumers whenever the social welfare penalties of supply interruption do not exceed the costs of supplying the service. Hence the French model incorporates these policy phases:

- periods when capacity matches output requirements.
- periods when there is surplus capacity.
- periods when capacity is less than required and selected blackouts are used to ration demand.

Both Australian and British electricity authorities however are charged with - above all - providing a continuous supply service so that in our models we can only expect to have the first two kinds of policy operating periods shown above.

The model we now examine therefore, while based on the French continuous time model so that we can examine the concept of permanent load increments, does not allow for all the policy options which that model considers.

The important exogenous variables remain the same as those of previous models we have discussed:

output requirements - which must be fulfilled: $X(t)$

running costs per unit of output: $r(X, t)$

capacity costs per unit of new capacity installed: $c(t)$

There are two important constraints; firstly output requirements must always be met, and we could either write this constraint explicitly or simply substitute, as we have done above, required output for actual output. Thus we will simply use the variable $X(t)$ for actual and required output. The second important constraint is that output cannot exceed capacity. The manner in which we have defined output as a continuous function of time implies that the unit of output is the kilowatt and this is also the unit of capacity. Hence the constraint is simply:

$$Q(t) \geqslant X(t)$$

where $Q(t)$ is capacity installed. As new investment carries on we are varying the amount of capacity available to us and we therefore have the simple definition: investment - which we shall call $u(t)$ - is the rate of change of capacity:

$$u(t) = \dot{Q}(t) = \frac{dQ(t)}{dt}$$

Investment, of course, can never be negative:

$$\dot{u}(t) \geq 0 \text{ for all } t$$

We have now the essence of a very simple investment planning model but it is sufficient to examine Turvey's concepts of permanent output increments because we can write a permanent output increment as a shift in the whole set of output requirements:

$$\Delta X(t)$$

and this will refer to a completely new output requirements programme beginning at a certain date and carrying on to the end of the planning horizon. For example $\Delta X(t)$ could be the difference between the two output programmes $X_1(t)$ and $X_2(t)$ shown in figure IX,2 above.

Now putting together the statement of the problem we can write:

minimize the present worth of system costs of meeting output requirements subject to the capacity restraint;

and stated mathematically this is

$$\text{minimize } C = \int_0^\infty \left[\bar{c}(t) \;\; u(t) + r(X,t) \right] e^{-it} dt$$

such that: $u(t) = \dot{Q}(t)$, $u(t) \geqslant 0$

$X(t) \leqslant Q(t)$

with given initial conditions, e.g. $X(0) = Q(0) = Z$.

We have introduced the discount factor e^{-it} with constant rate of discount i. In addition it is important to note that the price of new investment equipment, c(t), explicitly varies through time with the process of technological change; it is this dependence which ensures the model is truly dynamic; without it the solution to the model would not depend on time - in the language of dynamic optimization the problem would be "autonomous".

The appendix sets out the solution conditions of a dynamic optimization problem like this; here we will take these conditions as given and discuss their economic content. As usual the economist deals with a constrained optimization problem like this by redefining the objective function to include the constraints with each valued at a shadow price to be determined.

The most crucial part of the solution conditions will clearly be an investment rule. As we know from the statement of the problem investment can be either positive or zero

$$u(t) \geqslant 0$$

and the investment rule decides this. It is quite simple: suppose we write

$$\lambda(t)e^{-it}$$

as the Kuhn-Tucker multiplier associated with the capacity constraint:

$$X(t) \leqslant Q(t)$$

then following our usual analysis $\lambda(t)e^{-it}$ is the present worth in cost saving terms of an extra unit of capacity in year t of the optimal plan, i.e.

$$\frac{\Delta C^*}{\Delta Q(t)} = \lambda(t)e^{-it}$$

The cost of new investment is $c(t)e^{-it}$ per unit and such new investment is only undertaken if the cost savings are recouped over the unit of capacity's life. Thus

$$u(t) > 0 \quad \text{if} \quad c(t)e^{-it} = \int_t^\infty \lambda(y)e^{-iy}dy$$

where y is simply a dummy variable of integration. If this condition does not hold, i.e. if

$$c(t)e^{-it} > \int_t^\infty \lambda(y)e^{-iy}dy$$

then we have a boundary solution of zero investment for some time periods: $u(t) = 0$. Thus our optimal investment policy switches between periods when $u(t) > 0$ and periods when $u(t) = 0$. Figures IX, 4 and IX, 5 below illustrate this. In the upper figure (4) we show a possible output requirement path which at first rises then falls and subsequently rises again. The lower figure (5) shows what happens to capacity; at first

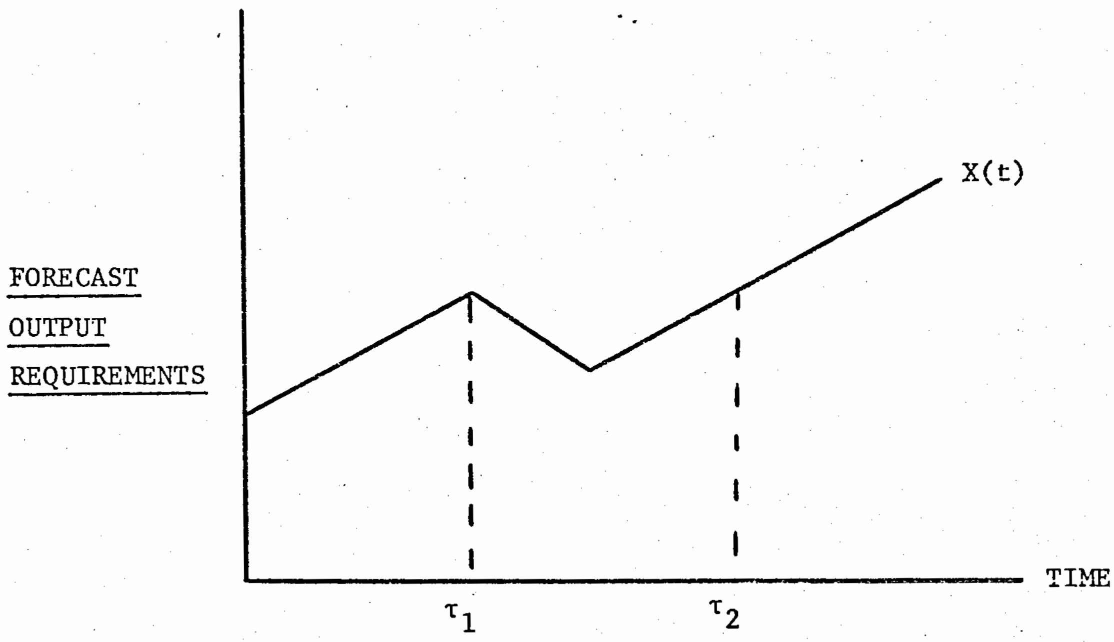


Figure IX,4

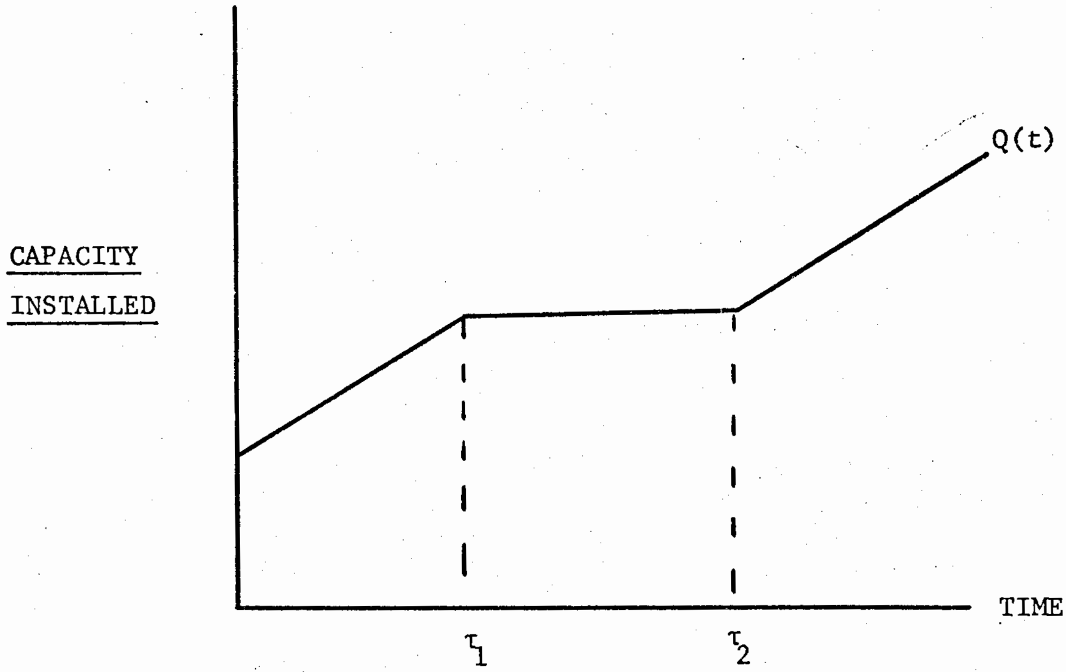


Figure IX,5

it is built up, then as the optimal policy switches to one of no investment, capacity remains static; eventually however it is again built up as positive investment takes over. From a cost viewpoint the instants of switching in the policy, τ_1 and τ_2, are of extreme importance. (The two points taken here are simply an example, although it is always the case that the number of switching points is finite for a feasible solution to the problem.)

There are some other solution conditions which provide economically meaningful results. The model provides an instantaneous measure of the term we have called λ: an instant's cost savings:

$$\lambda = (ic - \frac{dc}{dt})$$

This expression simply says that λ - an instants cost saving - is the sum of two terms: an interest cost element (ic) is the first term; the second (dc/dt) detracts from cost savings because it reflects the loss of technological progress in capacity costs implied by building capacity immediately instead of delaying for an instant.

Apart from providing an investment rule therefore, this model provides a simple expression for the cost of one instant's demand: a sum of a running cost element, r', and an instantaneous capital charge, λ.

Hence over the lifetime of the plan the costs of an infitesimal unit of demand are:

$$\int_{0}^{\tau_1} (\lambda+r')e^{-it}dt \quad + \quad \int_{\tau_1}^{\tau_2} r'e^{-it}dt \quad + \quad \int_{\tau_2}^{\infty} (\lambda+r')e^{-it}dt$$

The first and third present worth terms refer to periods $(0\tau_1)$ and $(\tau_2\infty)$ and are therefore associated with positive investment in capacity. During those periods demand requires that capacity be built and run. However during the period $(\tau_1\tau_2)$ when demand has slumped so that capacity is in excess, each small unit of demand is only responsible for a stream of running costs associated with existing capacity (the second term in the above expression).

This expression is the key to evaluating the cost of a permanent output increment as defined by Turvey. We want to shift the whole function X(t) and calculate

$$\frac{\Delta C^*}{\Delta X(t)}$$

The important factor is the <u>starting date</u> of the permanent output increment. Suppose we call it θ. θ could occur in any of the three periods

$$(0\tau_1) \; ; \; (\tau_1\tau_2) \; ; \; (\tau_2\infty)$$

If the permanent output increment occurs in the first or the third period we know that the optimal policy already requires positive investment to keep output up with capacity requirements; any output increment starting in these periods will require additional capacity to be built starting at date θ and incurring a stream of instantaneous extra system costs of:

$$\Delta C^*/\Delta X \quad = \int_{\theta}^{\infty} (\lambda + r')e^{-it}dt \qquad \text{..........} \; (1)$$

i.e. a stream of instantaneous 'marginal' capacity and running costs.

However if the output increment begins in the period of surplus capacity when investment is zero, $\tau_1 \tau_2$, then no extra capacity costs are incurred <u>until after τ_2</u> when output once more impinges on capacity limits, and then

$$\frac{\Delta C^*}{\Delta X} = \int_{\theta}^{\tau_2} r' e^{-it}dt \; + \; \int_{\tau_2}^{\infty} (\lambda + r')e^{-it}dt \qquad \text{.....} \; (2)$$

Hence the costs of a permanent output increment depend very much on when increment begins. These expressions for $\Delta C^*/\Delta X$, labelled (1) and (2), are what Albouy and Nachtigal would call "dynamic marginal cost". Turvey's 1969 concept is couched in terms of the postponement or bringing forward of the crucial starting date θ . Taking expression (1) first, the case where θ occurs in a period of capacity expansion, we calculate a variation in θ by postponing or advancing the date as

$$\frac{d}{d\theta} \; \frac{\Delta C^*}{\Delta X} \quad = \quad - \; (\lambda + r')e^{-i\theta}$$

thus there is both a capacity and a running charge.

However when we take the case of a variation in the starting date of an output increment that occurs in a period of surplus capacity we have:

$$\frac{d}{d\theta}\,\frac{\Delta C^*}{\Delta X} \;=\; -\,r'\,e^{-i\theta}$$

and here there is only a running charge variation.

This concept Albouy and Nachtigal refer to as "instantaneous dynamic marginal cost" and it is as close as we can get to Turvey's complex 1969 definition. The interesting conclusion that we get from the point of view of our previous analysis is that this definition involves us in looking at the sum of an instantaneous running charge and an instantaneous capacity charge and the capacity charge depends on the extent to which capacity constraints are binding on output possibilities.

Thus Turvey's interesting definition does not take us on a radically different road from the programming analysis we have already examined. This makes sense when we think out Turvey's analysis. An investment planning authority is forewarned of a permanent rise in output requirements and considers how to price it. The crucial question is when will the increment occur? Will it occur in a period of surplus capacity or in a period when capacity has to be expanded? If prices are to be changed to encourage or discourage this output increment they must reflect the date on which it begins. Turvey considers marginal cost to be a measure of the cost changes associated with variations in this all important date so his definition must capture the instantaneous costs of this output increment; naturally these will include a running charge and also a capacity charge which varies with the extent of the limitations on capacity. All this is exactly what the programming analysis does by a different route.

PRICING OF PERMANENT AND TEMPORARY OUTPUT CHANGES

Turvey raises several new issues in pricing polciy in his analysis and these same issues have also been discussed by the French economists. Most important is whether "dynamic marginal cost" or "instantaneous dynamic marginal cost" as we have defined them here is to be the basis of pricing policy. A new electricity consumer clearly does not expect to switch off his consumption after a short period; each new consumer brings along a permanent output increment. In this sense one would expect prices to be based on a measure of the costs of the permanent output increment.

The possible schemes can be illustrated quite easily using equations (1) and (2) from the previous section. Now we have two expressions for the cost of a permanent output increment

(a) the increment begins in a period of capacity expansion

$$\frac{\Delta C^*}{\Delta X} = \int_{\theta}^{\infty} (\lambda + r')e^{-it}dt = \Delta \quad (1)$$

(b) the increment begins in a period of surplus capacity

$$\frac{\Delta C^*}{\Delta X} = \int_{\theta}^{\tau_2} r' e^{-it}dt + \int_{\tau_2}^{\infty} (\lambda + r')e^{-it}dt = \Delta \ (2)$$

and (since λ and r' are both expected to be positively related to output) clearly

$$\Delta(1) \geqslant \Delta(2)$$

Thus charging a contract price equal to the present value sums $\Delta(1)$ or $\Delta(2)$ will encourage new customers to appear in periods of surplus capacity. But how will the consumer actually pay for his electricity; - so far we only require that for a long term supply he hand over the amount $\Delta(1)$ or $\Delta(2)$ according to his starting date. He may prefer to pay uniform installments, e.g. in the form of an annuity multiple of $\Delta(1)$ or $\Delta(2)$ (a situation which Turvey wished to avoid in his 1968 analysis - see the third section of this chapter):

$$i.\Delta(1) \quad \text{or} \quad i.\Delta(2)$$

This however does not differentiate between peak consumption periods and off peak consumption periods. As an alternative Albouy and Nachtigal suggested a "hybrid solution". The customer may elect to pay his investment charges in uniform installments: he then pays an annuity equivalent of either

$$\int_{\theta}^{\infty} \lambda e^{-it} dt \qquad \text{or} \qquad \int_{\tau_2}^{\infty} \lambda e^{-it} dt$$

and pays his running charges as they arise. This system of contractual payments still encourages the "arrival" of new permanent customers in periods of surplus capacity but regulates their subsequent peak consumption.

Turvey however dismisses the contractual payments approach in his 1971 treatment of pricing policy. He considers whether new customers should be faced with a long term schedule of marginal costs of supplying the required service but concludes that:

"... such contracts are only worth making with a few large customers ... "

(Turvey 1971a, page 57)

The basis for this is that such contracts for supply are not marginal "in the strict sense of relating to a tiny increment in output". He prefers to adopt the idea of instantaneous dynamic marginal cost to determine prices on offer to a large number of customers making short run (switching on or off) decisions.

It is rather paradoxical that having established the important idea of permanent output increments Turvey prefers to maintain exclusively an analysis which splits such increments up into an infinity of instantaneous increments. It would be preferable to regard both forms of dynamic marginal cost as having a part to play in pricing policy.

SUMMARY OF THESE RESULTS

We have covered a complex range of marginal cost definitions and can sum up as follows:

(1) we began with the programming approach and the dual variable nature of marginal cost.

(2) we introduced Turvey's 1969 definition which looked at marginal cost from an entirely new direction.

(3) this new approach makes the static concepts of short run and long run marginal cost irrelevant in a dynamic analysis. In particular, Kay's attack on Turvey's results was shown to be ill founded.

(4) it is possible to model the permanent output increment definition and our model was based on a recent French approach to the problem.

(5) dynamic economics throw up two marginal cost concepts; one related to permanent output increments and one related to instantaneous output increments arising from varying the starting date of a permanent output change.

(6) both concepts reduce to an analysis of running and capacity charges and the instantaneous dynamic marginal cost is exactly identical to the dual variable definition of marginal cost.

(7) either measure could be the basis of pricing policy according to the type of output increment under consideration.

MATHEMATICAL APPENDIX TO PART II

INTRODUCTION

This appendix covers the mathematical theorems of part II of the thesis. Some of these have also been used in the discussion of peak load pricing models in part I. There are four sections in the appendix. Section (i) describes the fundamental Kuhn-Tucker theorem which is at the heart of all constrained optimization techniques. Section (ii) is a brief application of the theorem to the problem of short run scheduling of given generating capacity to meet varying loads. This is the analysis referred to in chapter VI above. Section (iii) of the appendix sets out the modern techniques of optimal control theory and the calculus of variations and the special cases which are used in chapter IX. Section (iv) discusses briefly the essential dynamic nature of the solutions and the problem of stationarity and autonomous solutions mentioned in relation to Turvey's statement of the problem. Finally section (v) presents the explicit solution of the model of chapter IX.

(i) THE KUHN-TUCKER THEOREM

The nature of the problem tackled in this thesis by using the Kuhn-Tucker theorem can be expressed as:

$$\underset{x}{\text{Minimize}}\ F(x) \quad \text{subject to} \quad G(x) \geqslant b$$
$$\text{with } x \geqslant o$$

where we may be dealing with either scalars or vectors.

This is the most general formulation of the constrained optimization problem. The constraints: $G(x) \geqslant b$, where b is a set of constants, can include the special case of equality constraints which is tackled by classical Lagrangian methods; on the other hand linearity of the functions F(x) and G(x) makes this a linear programming problem.

The first step in Kuhn-Tucker analysis is to formulate a Lagrangian type function using a set of Kuhn-Tucker multipliers, λ

$$L\ (x,\ \lambda) = F(x) \ + \ \lambda\ (b - G\ (x))$$

Then, under certain conditions, a certain x^* solves the original constrained optimization problem if and only if there exits λ^* such that x^* and λ^* together provide a saddlepoint of the unconstrained function $L(x, \lambda)$ i.e. a minimum of L with respect to x and a maximum with respect to λ. The most important of these conditions is the constraint qualification condition which requires that there is some point in the feasible region of solutions where all the inequality constraints are satisfied as equalities. This is the condition referred to in the discussion of the Kuhn-Tucker multipliers as marginal costs in chapter VI above.

This theorem specifies nothing about the differentiability of the functions involved. However, where F(x) and G(x) are differentiable then the saddlepoint property is ensured by the necessary conditions:

$$\frac{\partial F}{\partial x} \geqslant o \; ; \; \frac{x \partial F}{\partial x} = o$$

$$\frac{\partial F}{\partial \lambda} \leqslant o \; ; \; \frac{\lambda \partial F}{\partial \lambda} = o$$

These can be recognised as the complementary slackness conditions used in the text.

(ii) SHORT RUN SCHEDULING OF GENERATING CAPACITY

This elementary version of a programming model is illustrated for completeness and as an example of the Kuhn-Tucker theorem just discussed.

There are several types (k) and vintages (v) of existing generating capacity: Q^{vk}. At any time t, output (O_t^{vk}) from each cannot exceed existing capacity:

$$O^{vk} \leqslant Q^{vk} \quad \text{for all v, k and t} \qquad \text{.......... (1)}$$

There are certain hourly output requirements X_t which must be met by summing output from all vintages and types of capacity:

$$\sum_v \sum_k O_t^{vk} \geqslant X_t \quad \text{for all t} \qquad \text{.......... (2)}$$

Output can never be negative:

$$O_t^{vk} \geqslant o \quad \text{for all v, k, t} \qquad \text{.......... (3)}$$

Subject to the restraints (1) (2) and (3), the objective is to minimize total system running costs in this short run scheduling operation. If r_t^{vk} is the running cost of vintage v, type k capacity in hour t, then the Lagrangian of the problem is simply

$$\text{Minimize } L = \sum_t \sum_v \sum_k r_t^{vk} O_t^{vk} + \sum_t \sum_v \sum_k \lambda_t^{vk} (O_t^{vk} - Q^{vk})$$

$$+ \sum_t \mu_t (X_t - \sum_v \sum_k O_t^{vk}) \qquad \text{.......... (4)}$$

According to the Kuhn-Tucker theorem, we now have to find the set of values O_t^{vk}, λ_t^{vk} and μ_t that provide a saddlepoint of L. Assuming differentiability, the necessary conditions are:

$$\frac{\partial L}{\partial O_t^{vk}} \geqslant o \; ; \quad O_t^{vk} \cdot \frac{\partial L}{\partial O_t^{vk}} = o \qquad \text{.......... (5)}$$

$$\frac{\partial L}{\partial \lambda_t^{vk}} \leqslant o \; ; \quad \lambda_t^{vk} \frac{\partial L}{\partial \lambda_t^{vk}} = o \qquad \text{.......... (6)}$$

$$\frac{\partial L}{\partial \mu_t} \leqslant o \; ; \quad \mu_t \frac{\partial L}{\partial \mu_t} = o \qquad \text{.......... (7)}$$

Using condition (5), the "complementary slackness" condition implies that for any $O_t^{vk} > o$ – i.e. for any plant actually in production at hour t

$$\lambda_t^{vk} = \mu_t - r_t^{vk}$$

μ_t (= $\Delta C^*/\Delta X_t$) is, as usual, the marginal generating cost of the system. λ_t^{vk} is the Kuhn-Tucker multiplier reflecting the cost saving associated with each plant in each hour. It is the difference between marginal generating cost of the plants in operation and a particular plant's running cost in that hour. Plants for which $\lambda_t^{vk} < o$ will not be operated. Plant will be scheduled into operation in the order in which r_t^{vk} approaches μ_t. (Of course μ_t itself varies with each hour and with the different plants in operation.) Thus a schedule (in descending order) of the variables λ_t^{vk} is identical to the merit-order schedule of generating plant to be brought into operation.

(iii) <u>OPTIMAL CONTROL AND THE CALCULUS OF VARIATIONS</u>

Optimal control theory is the outcome of recent advances in the classical problems of calculus of variations. Although most dynamic optimization problems can be treated analytically using the Kuhn-Tucker programming approach, optimal control theory allows for a more explicit treatment of the truly dynamic properties of the solution. The problem is formulated in terms of an objective <u>functional</u> rather than a function, so that instead of solving for optimal values of the variables, one solves for optimal time functions of the variables.

The technique usually distinguishes between a <u>state variable</u> whose path one wishes to control in order to maximize a given measure of benefit and a <u>control variable</u> whose path controls the path of the state variable.

The objective functional will be some measure of benefit measured perhaps as a sum of all the benefits obtainable at each instant on the time path of the state and control variables. Thus one may seek to

$$\text{Maximize } J = \int_{o}^{T} F\ (x,\ u,\ t)\ dt$$

where $x = x(t)$ is the state variable

$u = u(t)$ is the control variable

and $t =$ time.

One wishes to choose $u(t)$ to control $x(t)$ so that J is a maximum. This is applied to some dynamic system whose evolution is governed by the differential equation:

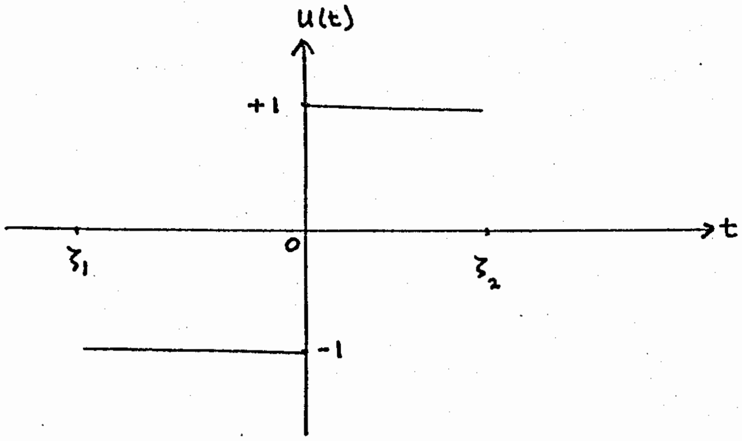
$$\dot{x} = \frac{dx}{dt} = f\ (x,\ u,\ t)\ ;\quad x(o) = x_o$$

Optimal control theory is flexible enough to treat cases where there are constraints operating on the available choice of $u(t)$. However it does require that $u(t)$ is a piecewise continuous function of time, t. As an example consider the function

$$u(t) = -1 \qquad \zeta_1 < t < o$$

$$u(t) = +1 \qquad o < t < \zeta_2$$

which graphs as



This ensures that there are a finite number of "jumps" for u(t) such as the one occurring at t=o. The control variable <u>switches</u> from one path to another. Putting this together, the optimal control problem is

$$\text{Maximize by choice of } u(t) \quad J = \int_o^T F\,(x,\ u,\ t)\ dt\,,$$

$$\text{such that} \quad \dot{x} = f\,(x,\ u,\ t)$$

$$x(o) = X_o$$

u(t) subject to certain restraints.

There are a variety of conditions governing the existence of a solution, but of chief interest here are the analytical conditions defining the solution.

The method is to set up a function incorporating the instantaneous value of the objective functional and the basic differential equation of the system: the Hamiltonian function

$$H(x, u, t, p) = F(x, u, t) + pf(x, u, t)$$

where $p = p(t)$ is known as a co-state variable. The solution is found if $p(t)$ exists such that the following conditions hold:

(a) $\dot{p} = \dfrac{-\partial H}{\partial x}$

(b) $\dot{x} = \dfrac{\partial H}{\partial p}$

These two differential equations define the optimal dynamic system. One set of boundary conditions is given by $x(o) = x_o$. There may also be the requirement $x(T) = x_T$; otherwise some transversality conditions can be developed:

$$p(T) \geqslant o \quad ; \quad x(T)\,p(T) = o$$

Finally the crucial choice of $u(t)$ to be used in solving (a) and (b) comes from solving:

(c) Maximize $H(x, u, t, p)$ with respect to $u(t)$. (These results are set out and proved in Arrow and Kurz (1971)(pages 33-57.)

Difficulties arise when $T = \infty$, i.e. we deal with an infinite horizon. There are no longer necessary boundary conditions though sufficient conditions are (again see Arrow and Kurz (1971) page 46)

$$\lim_{t\to\infty} p(t) \geqslant o \quad ; \quad \lim_{t\to\infty} p(t)\, x(t) = o$$

A special case arises when there are constraints on the state variable x(t); e.g.

$$g\ (x,\ u,\ t) = o$$

Now, in addition to the Hamiltonian function, a Lagrangian is also constructed

$$L\ (x,\ u,\ t,\ p,\ q) = H\ (x,\ u,\ t,\ p) + q\ g\ (x,\ u,\ t)$$

Condition (c) remains unchanged but (a) and (b) become

$$(a^1) \quad \dot{p} = \frac{-\partial L}{\partial x}$$

$$(b^1) \quad \dot{x} = \frac{\partial L}{\partial p}$$

and the additional constraints must be satisfied.

These theorems are applied to the marginal costing problem in section (v) of this appendix, below. This section is the basis of the results of chapter IX.

There is an alternative, more traditional approach to problems of dynamic optimization in the calculus of variations. This treats problems of the nature:

$$\text{maximize} \quad I = \int_o^T \emptyset\ (x,\ \dot{x},\ t)\ dt$$

(Simply defining $\dot{x} = u$ turns this into an optimal control problem.) The necessary conditions are given by the second order Euler differential equation:

$$\frac{\partial \emptyset}{\partial x} - \frac{\partial d}{dt} \frac{\partial \emptyset}{\partial \dot{x}} = o$$

If however $\dot{x}$ did not enter the objective functional I in a meaningful way, then

$$\frac{\partial \emptyset}{\partial \dot{x}} = o$$

and the Euler equation becomes simply

$$\frac{\partial \emptyset}{\partial x} = o$$

which is not a differential equation at all, but simply reduces to a static condition to be repetitively met each instant of time. This is the essential criticism of the Jorgenson neoclassical investment model described in chapter VII above. Jorgenson's crucial variable is the optimal value of the capital stock, K(t), but the rate of growth of this stock - although it affects output by augmenting "undecayed" capital equipment - has no effect on profits or costs. It can be substituted out and the Euler equation reduced to the static case. The Gould-Lucas-Nerlove critiques referred to in chapter VII's discussion all associated positive "adjustment costs" with variations in $\dot{K}(t)$ and hence ensured that the optimality conditions were differential equations, thus maintaining the truly dynamic character of the problem.

(iv) A SPECIAL CASE OF THE OPTIMAL CONTROL PROBLEM

This comment is really a development of the comment at the end of the last section. Chapter VI emphasised that Turvey had taken some pains to maintain the dynamic nature of his solution by explicitly incorporating technical progress, (and hence time, t) into his model. This can be shown by using the optimal control theorem described above. Suppose that the objective measure of benefit and the basic differential equation of the system do not depend explicitly on time

$$F(x, u, t) = F(x, u)$$

$$f(x, u, t) = f(x, u)$$

then there may be a solution where the following hold (Arrow and Kurz (1971) page 50):

(i) the value of the Hamiltonian function is constant through time

(ii) the differential equations for $\dot{x}$ and $\dot{p}$ are <u>autonomous</u>.

In this case certain boundary conditions (starting and finishing points) will provide a long run <u>stationary</u> equilibrium solution with $\dot{x} = \dot{p} = o$ - i.e. all dynamic motion ceases. This is exactly the mathematical analogy of the static solution of the text-books which Turvey wishes to avoid. Hence to avoid the possibility of <u>stationary</u> solutions it is essential that time, t, explicitly affects the costs and benefits of the exercise and this can only be achieved by specifically allowing for <u>technical progress</u> in calculating costs.

(v) <u>THE DYNAMIC MARGINAL COST PROBLEM</u>

Only the essential economics of the problem is presented in chapter IX above and the mathematical background is set out here.

The problem as stated in chapter IX is:

$$\text{minimize } C = \int_{o}^{\infty} \left[c(t)\, u(t) + r(X, t) \right] e^{-it}\, dt$$

$$\text{such that } u(t) = \dot{Q}(t),\ u(t) \geqslant o$$

$$X(t) \leqslant Q(t)$$

$$X(o) = Z$$

In this case the basic differential equation of the system is simply $u(t) = \dot{Q}(t)$ and $u(t)$ is constrained not to be negative. The problem is not autonomous, since both running and capacity cost vary with time.

This is an optimal control problem with infinite horizon and constraints on the state variable $X(t)$. To solve, set up a Hamiltonian function

$$H\ (X,u,t,p) \quad = \quad -e^{-it}\ \left[c(t)\ u(t) + r(X,t)\right] + p(t)\ u(t)$$

and a Lagrangian function

$$L\ (X,u,t,p,\ ) = H\ (X,u,t,p) + \lambda e^{-it}\ (Q - X)$$

(Note, the minus sign appears before the cost terms because the problem is one of minimization.)

Now apply the conditions of the optimal control theorems, beginning with the maximization of $H(X,u,t,p)$ with respect to $u(t)$. This requires

$$\frac{\partial H}{\partial u} \leqslant o$$

If $\partial H/\partial u = o$ there is an interior solution with $u(t) > o$ and if $\partial H/\partial u < o$, then the solution requires $u(t) = o$

$$\frac{\partial H}{\partial u} = -c(t)e^{-it} + p(t)$$

Thus u(t) switches from zero to positive according to whether $c(t)e^{-it}$ exceeds or equals p(t). To determine p(t) we use the differential equation condition

$$\dot{p}(t) = \frac{\partial L}{\partial Q(t)}$$

i.e. $$\dot{p}(t) = \lambda(t)e^{-it}$$

and integrating

$$p(t) = \int_t^\infty \lambda(y)e^{-iy}\,dy$$

where y is the dummy variable of integration. Hence u(t) is positive when

$$c(t)e^{-it} = \int_t^\infty \lambda(y)e^{-iy}\,dy$$

and u(t) = o when

$$c(t)e^{-it} > \int_t^\infty \lambda(y)e^{-iy}\,dy$$

These are the investment rules presented in the text. u(t) will switch from positive to zero and back again according to which of these conditions holds at each point of time.

PART III

ESTIMATING THE DEMAND FOR ELECTRICITY

CHAPTER X

THE RESIDENTIAL DEMAND FOR ELECTRICITY AT THE STATE LEVEL

INTRODUCTION

The purpose of the next three chapters is to investigate the demand for electricity in the residential sector by fitting demand equations both nationally and at the state level. The residential sector is simply defined as those consumers who are metered on domestic or residential tariffs as opposed to those metered on commercial and industrial tariffs. This categorisation is well defined today though in the immediate post-war period some consumers were put into "unclassified" categories because they were metered on special forms of tariff. This problem however had effectively disappeared by the beginning of the longest of the sample periods used here, June 1953.

It is a reasonable approach to consider the residential demand for electricity separately from industrial and commercial demands, for two chief reasons. Firstly the utilities themselves consider these demands as entirely separate and approach the forecasting problems of each in different ways. Secondly the development of demand theory has distinguished between consumer demand for final products and producer demand for intermediate goods. Consequently the application of consumer demand models will proceed much more readily in the case of residential electricity demand.

One logical way to carry out the investigation of residential electricity demand in a country like Australia would be as follows. Beginning at the State level, the level at which most electricity consumption data is collected and published, one can attempt to fit the conventional models of demand analysis, extending these wherever possible in the directions especially applicable to energy demand. Sooner or later, however, one is faced with two serious problems. Firstly, conventional models of energy demand skirt around the problem of relating that demand to the demand for the appliances which are complementary to the use of any given fuel. It becomes apparent that one must bring into the analysis of electricity demand some consideration of the factors influencing consumer durable demand in a manner more direct than has so far appeared in the literature[1]. Consequently, the number of explanatory variables that can be used in explaining electricity demand rises considerably when explicit recognition is taken of the complementary demand for consumer durables.

The second problem is that, at the State level, the sample period is relatively short. Effective classification of residential demand begins in some States (e.g. Victoria and N.S.W.) in the early 50's and is not adopted by others (e.g. Queensland) till the early 60's. As a result the samples using annually collected data can have at the most about 19 observations and frequently less. Both these problems - the large number of explanatory variables and the small sample size at the State level - mean that one must eventually turn towards estimating a national demand equation using pooled samples of cross section and time series data. This effectively solves the degrees of freedom problem that especially arises when one wants to take explicit account of the demand for consumer durables. The sample size used in the national equations is greater than those used in most of the well-known macroeconometric models in use in Australia. However, it does mean that another data problem has to be faced, that of non-homogeneity of subsets of the sample when it is pooled from time series and cross section observations. It will be seen below how this part of the study is tackled. To sum up therefore, the investigation begins with time series equations at the State level, which are of prime interest to the suppliers of electricity, and when the degrees of freedom problem becomes apparent, moves over to pooled sample equations at the national level. Other aspects of the data problems will become clear as the study proceeds.

PRIOR CONSIDERATIONS

Before we can commence the estimation of the demand equations the prior question of the structure of the economic model we are looking at has to be considered. The outstanding prior consideration in setting up demand equations is of course the problem that arises from the possibility that the fitted equation may be a linear combination of a demand equation and a supply response. However, consideration of the institutional characteristics of the economic setting can often overcome this and this has been particularly shown to be the case with energy demands and supplies.

Almost every form of energy production is highly capital-intensive and none more so than electricity production. As a result there are very long lags in the response of supply to any new developments in demand characteristics, once a given generating system has been installed.

It is important here to distinguish between instantaneous load manipulation and overall load growth. If we investigate the demand for electricity as it develops during one 24 hour period we can see supply responding virtually instantaneously, as the unit of output is an additional kWh of already-installed generating capacity up to the point of effective full capacity beyond which "load shedding" begins. In this case supply response to demand is all important and a form of price rationing mechanism between supply and demand must be explicitly used. Hence a simultaneous equations model was used in chapter IV's treatment of peak load pricing. However if we investigate, as we are doing here, the annual growth of demand in some particular sector, the responsiveness of supply is completely different. The unit of supply is now an extra amount of generating capacity and the provision of additional generating capacity requires very long gestation periods. For instance, in the Victorian State Electricity Commission an example of a new unit of supply is the additional power station project at Newport which was authorized in 1971 and is expected to be complete in all stages (1st and 2nd generating units) by 1978[2]. Smaller units of supply could be adopted but their effective responsiveness to demand developments is very lagged. This factor makes it essential, in annual demand studies, to consider the sort of model that is illustrated by the following elementary example

$$Q_t = a P_t + u_t \qquad \ldots\ldots\ldots\ldots \qquad (1)$$

$$P_t = b S_t + v_t \qquad \ldots\ldots\ldots\ldots \qquad (2)$$

$$S_t = \sum_{i=1}^{i=n} c_i Q'_{t-i} + w_t \qquad \ldots\ldots\ldots\ldots \qquad (3)$$

Here, Q_t is consumption at time t, P_t is price at time t, S_t is installed generating capacity at time t, and u_t, v_t and w_t are random disturbance terms. This model states that current consumption depends on a set of prices that are fixed in the current period (equation (1)). This is a simple representation of the no-bargaining position between electricity utilities and consumers. The utility sets a block of tariffs which, for current consumption, the consumer must take as given. This set of tariffs will usually reflect marginal costs of producing electricity from currently installed generating capacity and will usually include some rationing charge for peak-load demands. For the individual residential consumer there is no possibility of bargaining or haggling over this tariff and in large measure this applies to industrial and commercial consumers as well. Although the latter have the possibility of negotiating "interruptible" supply contracts, whereby at times of system peak demand their supplies are cut off and they automatically switch to a standby fuel, (usually oil or gas) for which they receive tariff discounts, nevertheless this possibility cannot be regarded, at least in Australia, as providing an opportunity for large scale bargaining over current tariffs[3].

Turning now to the second and third equations, these show that the tariff structure is usually set up on the basis of the costs of existing generating capacity, which will usually have a fairly high average age given the highly capital intensive nature of electricity production. Even if a new unit of output was coming on stream each year, current tariffs can at most be based on a supply capacity that is at least one year old. Equation (2) reflects this formulation of price determination. Equation (3) then shows that currently installed generating capacity will be some (perhaps complex) function of past consumption. The only currently dated variable that might enter the right hand side of equation (3) is a forecast of current consumption Q_t made from previous consumption data. <u>However actual current consumption cannot influence current generating capacity.</u> Hence equations (1), (2) and (3) reflect two outstanding characteristics of electricity demand and supply: the lack of bargaining over current tariffs and the lag in the responsiveness to consumption changes.

The third equation contains only predetermined variables on the right hand side and hence can be separated from the model. So that equations (1) and (2) can now be rewritten in the familiar structural form:

$$\begin{bmatrix} 1 & 0 & -b \\ -a & 1 & 0 \end{bmatrix} \begin{bmatrix} P_t \\ Q_t \\ S_t \end{bmatrix} = \begin{bmatrix} v_t \\ u_t \end{bmatrix}$$

and the matrix of coefficients of the jointly dependent variables, P_t and Q_t is therefore triangular. If we assume in addition that the random disturbances operating on current demand, u_t, are independent of the random disturbances, v_t, operating on the pricing of supply from existing capacity, then the matrix of variances and co-variances of the disturbances:

$$\begin{bmatrix} Eu_t^2 & Eu_tv_t \\ Ev_tu_t & Ev_t^2 \end{bmatrix} \quad \text{is diagonal.}$$

As a result of these manipulations, the simple model set up to reflect the leading characteristics of electricity supply and demand is apparently recursive and hence both equations are just identified (each equation is equivalent to a reduced form equation). More importantly, ordinary least squares provides unbiased estimates of the structural coefficients.

The models we consider in representing this demand for electricity are more complex than the equations above illustrate. Nevertheless the institutional characteristics shown above remain one of the most obvious explanatory variables in the demand equation, price, is exogeneous in the current time period for almost all consumers and the responsiveness of supply is very lagged - with lags at least as great and probably greater than the period between annual observations.

To sum up these considerations prior to estimation: economic models of electricity demand must pay attention to the fact that price is exogenous in the demand equation and current supply is not responsive to price. Accepting this means we can sensibly use ordinary least squares as an estimation technique and at the same time be sure we are estimating true demand relationships and not linear combinations of demand and supply responses.

DATA ON ELECTRICITY IN AUSTRALIA

All the electricity data in this chapter are obtained from annual reports of the Electricity Supply Association of Australia. These publish figures on consumption of electricity in each State by summing the sales of individual utilities within each State. However, the data exclude sales in the Australian Capital Territory and the Northern Territory (ACT and NT) as these are not consistently reported over the sample periods considered. Furthermore, not all States report their data on the same basis, and this has had some relevance for the choice of sample period.

Consistent reporting of electricity consumption in each state begins in the early 1950's and therefore the starting period of observations is the year July 1st 1953 to June 30th 1954. All data are reported in terms of the financial year, which, unfortunately, is a poor choice of observation period for energy data. Given Australian geographical conditions, the time of maximum energy demand often falls just at the point when one observation period ends and another begins. This has the effect of taking, say, a random weather disturbance and distributing it equally over two successive consumption observations, thereby reducing the variance of the data on consumption. For example suppose the months of June and July are particularly cold one year;

electricity consumption should show a particularly large deviation above trend. However this deviation is split so that part falls into the observation period ending June 30th and part in the observation period beginning July 1st and the data only reveal two small successive deviations above trend. The overall variance of the data is reduced, though the movement of successive deviations in the same direction introduces an element of serial correlation into the regressions. Since in the regressions at the state level weather is treated as a random disturbance, this may affect the interpretation of some of the results. There is some evidence, in the form of the Durbin-Watson statistic, that this may be occurring, though not seriously and not especially in the warmer states where it might be expected.

New South Wales, Victoria and South Australia have the longest history of publication of data separately on residential and industrial electricity consumption, yielding 18 annual observations in each State over the period 1953-54 to 1970-71. West Australia comes next with 17 annual observations and then Tasmania, the most important state in terms of electricity consumption per head, with 16 annual observations. Queensland has only published separate estimates of residential and industrial sales since the early 1960's and gives only 8 annual observations. Some idea of the relative intensity of electricity use in each state is given by the following table.

Table X 1. Average residential electricity consumption per head: 1962-3 to 1970-1 (KWhr)

NSW	VIC	QSLD	SA	WA	TAS
1140.4	905.6	884.2	1034.6	627.6	1886.7

The well-known aspects of received demand and utility theory are of course expressed in terms of the behaviour of the individual consumer, and for this reason demand equations are conventionally expressed in per capita terms, at least for consumers. Of course this only scratches the surface of the aggregation problem and, in particular, expressing demand data in per capita terms involves giving equal weight to the behaviour of each customer. To the extent that this is a violation of true behaviour patterns, we must regard the unweighted average of electricity consumption as a biased estimate of the true weighted average. There are additional reasons however for using per capita data. Clearly population need not then be included as a separate explanatory variable; thereby removing a possible source of multi-collinearity although there are models that have been used in per capita terms where population has been called on to provide explicit explanation of economic phenomena[4]. In addition it has been pointed out (Houthakker and Taylor (1970)) that demand equations regressed in per capita terms often show greater stability in the estimated relationships than aggregated demand equations. Bearing these considerations in mind, consumption income and durable stocks data are expressed in per capita terms.

Besides electricity consumption, the other important electricity variable is price, and several comments are in order on the use of this variable. Electricity is normally sold on some sort of two-part tariff involving a standing charge to recover capacity costs by appropriating some of the consumer surplus from electricity usage and a running charge, to reflect the cost of the marginal kWhr. Houthakker has pointed out[5] that at the margin choices of whether or not to take on extra electricity consumption will be sensitive to the running charge but not the energy charge. Ideally we should like an index of marginal electricity price to residential consumers at the state level. At the time when this demand study was carried out this was not available and enquiries to the Department of National Development on this matter reveal that even there average revenue is the only usable estimate of electricity price. Hence this is the estimator used, and it should be borne in mind that it contains an element (the energy charge) which does not apparently influence the demand for electricity at the margin. On the other hand it could be argued that the standing charge is an important factor in determining whether new consumers adopt electricity as a source of energy at all[6]. However this may no longer be a very relevant choice for many consumers except in some rural areas[7].

Unfortunately this is not the only load that the average revenue data has to bear. For reasons which are apparently confidential, no data is available on gas sales or prices at the state level in most states of Australia. The policy of the National Gas Association was that all such data after 1961 was not to be generally available at the time of writing. As a result the competitive influences of gas sales can only be revealed in our analysis by absorbing them into electricity prices, as the following example illustrates. Suppose there is a reduction in the price of electricity and a simultaneous rise in the price of gas, then any rise in electricity consumption will be due to a combination of the two but in the fitted equations will be apparently all due to the fall in electricity price. Hence the apparent own-price elasticity of electricity demand will be over-estimated. Of course this effect will not be so serious if gas prices and electricity prices move in step in the same direction. Certainly electricity prices have fallen steadily in every State over the period of the 1950's and 1960's. With gas prices we cannot be certain what has happened except to say that the recent introduction of natural gas in some states had definitely resulted in falling gas prices. These points should be borne in mind when we consider the work done by the electricity price variable.

OTHER VARIABLES

All the variables expressed in monetary terms are deflated by the consumer price index operating in the capital city of the state in question. The income variable used is personal income per capita deflated by the consumer price index, and modified to exclude the influence of the ACT and NT. All the data apart from the electricity figures are published by the Commonwealth Statistician. As more complex variables are introduced they will be discussed.

STATIC DEMAND MODELS

As we saw above, the data are used in terms of the "average" or "representative" consumer at the State level. In accepted utility theory, an individual consumer maximizes his utility flow from the amounts of goods he consumes subject to a restraint on his income or wealth. The equilibrium conditions of the consumer's behaviour then allow us to solve for the demand schedule for the jth good in the utility function as:

$$Q_j = f(P'_i \ldots P'_n, Y) \qquad \ldots\ldots\ldots\ldots (4)$$

where Q_j is the demand for the jth good, $P'_i \ldots P'_n$ are the prices of all the goods entering the utility function and Y is the consumer's income or wealth. In econometric studies we normally adopt a numeraire with which to measure relative prices of goods and a convenient numeraire is a weighted average of all prices: then P_j where

$$P_j = P'_j / \sum_i^n w_i P'_i$$ (and where the denominator is a weighted average price index)

is a convenient measure of the relative price of the jth good that removes some of the necessity for having prices of substitutes in the demand equation. Thus the usual fitted demand equation is:

$$Q_j = f(P_j, Y) \qquad \ldots\ldots\ldots\ldots (5)$$

Equation (5) can be fitted in several forms but two of the most convenient are the linear form (6) and the double logarithmic form (7).

$$Q_j = a + bP_j + cY \qquad \text{..........} \; (6)$$

$$Q_j = aP_j^b \, Y^c \qquad \text{..........} \; (7)$$

Equation (6) is the obvious form for regression purposes but has the drawback that the own-price elasticity can be anywhere in the interval $(- \infty, o)$, and usually some convention has to be adopted about where to measure the elasticity. In computational terms, the most convenient point is at the respective means of each variable. Equation (7) has proved more convenient for many demand studies because the regression coefficients in the logarithmic version, (8), are estimates of the respective demand elasticities:

$$\ln Q_j = \ln a + b \ln P_j + c \ln Y \qquad \text{..........} \; (8)$$

Both types of formulation are used here, although each implies different assumptions about the distribution of the random disturbances.

In essence this is a static model of electricity demand, since it considers electricity as a final good with no important complements. In particular it leaves out of consideration the relationship between the demand for electricity and the demand for consumer durables for which electricity is an input. However, this purely static model is a convenient starting point for the analysis, since some economists feel it reflects what might be called the "short run" demand for electricity; in other words the demand that would appear if all holdings of consumer durable stocks were held constant[8].

Tables X 2 and X 3 report the results from fitting equations (6) and (8) to electricity consumption per head in the six Australian states. Table X(2) gives the elasticities for the linear model computed at the means and table X(3) gives the estimates of the constant elasticity demand equation; (in other words the actual regression coefficients from equation (8)).

As it turns out this naive model appears to fit the date quite well on the conventional criteria. For the linear version 11 out of the 12 regression coefficients have the expected a priori sign and for the double log-version 10 of the 12 coefficients have the right sign. As might be expected personal income appears to be an important determinant of electricity consumption but at the same time there appears to be some inconsistency between the two sets of estimates. In the linear version the estimated income elasticity appears to be greater than one in every state except Queensland (where it is 0.98). This suggests that electricity demand is rather sensitive to income changes. If we accept this, it is clear that some intervening mechanism must be operating since it is hard to imagine a consumer raising his electricity consumption when his income rises without first buying some new form of consumer durable. Usage of already owned durables will not be particularly sensitive to income changes. Clearly then the linear version suggests the necessity of examining the link with the demand for durables. In this context the logarithmic version of the simple static model makes more sense, although the results do not appear to be as good

The column labelled D.W. reports the Durbin Watson Statistic.

Table X 2. Static demand model: Linear version

STATE	PRICE ELASTICITY	INCOME ELASTICITY	R^2	S	D.W.
NSW	-0.97**	1.32**	0.98	1574.6	2.03
VIC	-1.51**	1.14**	0.98	1365.6	1.34
QSLD	-1.22**	1.12**	0.97	1067.9	1.36
SA	-1.28**	0.98**	0.99	1282.2	0.99
WA	-0.69**	1.47**	0.98	688.3	1.27
TAS	0.3	1.12**	0.88	12321.3	1.47

note: ** denotes that the corresponding regression coefficient is at least twice its standard error and * denotes that the coefficient is greater than its standard error.

Table X 3. Static demand model: Double log version

STATE	PRICE ELASTICITY		INCOME ELASTICITY		R^2	S	D.W.
NSW	-1.62	(0.43)	0.76	(0.43)	0.96	0.0069	1.00
VIC	-2.58	(0.42)	0.45	(0.32)	0.96	0.0046	1.09
QSLD	-1.42	(0.35)	0.97	(0.32)	0.97	0.0015	1.31
SA	-2.28	(0.15)	-0.45	(0.23)	0.99	0.0016	0.90
WA	-1.64	(0.18)	0.48	(0.20)	0.99	0.0027	2.12
TAS	0.38	(0.48)	1.125	(0.15)	0.85	0.0053	1.32

note: figures in brackets after elasticity estimates are the standard errors.

as those of the linear model. In the logarithmic version only Tasmania shows an income elasticity greater than one. Apart from the odd result for South Australia, the other states show income elasticities lying between 0.45 and 0.76 which perhaps makes more a priori sense if the simple static model is to be regarded as a reasonable representation of true behaviour patterns.

Turning now to the price variable, the evidence seems to point to a highly competitive structure in the residential energy market. Leaving aside Tasmania where price seems to have no influence, the price coefficients are everywhere greater than twice their standard errors, and the estimated price elasticity is greater than one in 8 out of the 10 remaining cases. While electricity consumption has grown over the period studied, electricity tariffs have been falling, presumably for at least two reasons: technical progress in electricity generation and competition from gas, especially natural gas. Clearly the price elasticity estimates here are picking up a great deal of these competitive influences and we should recall at this point the comments made earlier about the absorbtion of gas price influences into own-price elasticities. If the demand for electricity is really as price sensitive as these estimates seem to indicate then, clearly, the tariff-setting policies of the electricity utilities are going to be fairly important in social welfare terms. The exception to these results is of course Tasmania, a state with a long history of electricity usage because of the low cost supply of hydroelectrical power. The demand and supply situation in Tasmania can be expected to be very different from that obtained in other states in Australia for several reasons. Firstly there is little competition, at least in the residential sector, from other forms of fuel so that the opportunities for consumers to substitute between different types of energy is very limited. Secondly, opportunities for technical progress in hydroelectricity generation may not be as plentiful as in the generation of electricity as a secondary form of energy, so that there may have been less opportunity for tariff reductions in Tasmania.

We may sum up this section by saying that while the naive model fits quite well, it suggests that we might profitably look at "longer run" formulations.

DEMAND MODELS WITH TIME TRENDS

The point has been made several times (and was emphasised by the above results) that the demand for electricity has to be related to the demand for electricity using consumer durables. We may hypothesise that a consumer will demand a certain flow of services from different durables and his demand for this flow of services will determine demand for electricity to use as an input in producing the services from his consumer durables. This section and those following attempt to apply this type of hypothesis to the demand for electricity, beginning with a very simple model, though one heavily relied on in the past,[9] and extending it to include more complex explanatory influences.

If we let W_t be the individual consumer's holdings of a stock of electricity-using durables then his demand for electricity could be written as:

$$Q_t = k_t W_t \qquad \text{.......... (9)}$$

where k_t is rate of appliance utilization at time t. The obvious determinants of k_t might then be the sort of economic variables we have been considering so far. Thus we may hypothesize:

$$k_t = aP_t^b Y_t^c \qquad \text{..........} \quad (10)$$

so that the demand schedule for electricity becomes:

$$Q_t = aP_t^b Y_t^c W_t \qquad \text{..........} \quad (11)$$

One implication of this precise formulation is that the elasticity of electricity demand with respect to durable stocks is one: a 1% rise in durable stocks calls forth a 1% rise in electricity demand, when price and income remain constant. This point can be reconsidered when we examine our regression estimates.

If equation (11) is to be used as our demand model we either require direct observations on consumer durable stocks or some form of equation to determine W_t. In our case we have to rely on the latter, although, in any case, data on holdings of durables has been notoriously deficient wherever it has been used.

There have been several noteworthy attempts to derive equations for variables like W_t, usually based on some kind of trend curve incorporating a saturation point. As a working approximation to a more sophisticated growth curve we could hypothesize the simple exponential:

$$W_t = W_o e^{rt} \qquad \text{..........} \; (12)$$

where r is the instantaneous rate of growth of electricity-using consumer durables. Combining equations (11) and (12) and writing aW_o as B we have:

$$Q_t = BP_t^b \, Y_c^c \, e^{rt} \qquad \text{..........} \; (13)$$

and this simple model then incorporates into the demand for electricity some allowance for the demand for the services of a stock of consumer durables. Rewriting in the form of natural logarithms gives our regression equation:

$$\ln Q_t = \ln B + b \ln P_t + c \ln Y_t + rt \qquad \text{..........} \; (14)$$

so that again regression coefficients are estimates of elasticities and the coefficient on the time trend is an estimate of the rate of growth of electricity using durables. Table X 4 presents the results for fitting this model to the sample periods described above. (For computational reasons the model was actually fitted in first difference terms: so that r was estimated as the intercept term. This has the effect of removing some serial correlation from the residuals.)

Table X 4. Demand model with time trend.

STATE	PRICE ELASTICITY	INCOME ELASTICITY	r%
NSW	-0.20	0.10	6.45
	(0.20)	(0.28)	(1.28)
VIC	-0.20	0.17	5.64
	(0.31)	(0.26)	(1.26)
QSLD	-0.27	-0.07	5.85
	(0.24)	(0.19)	(1.25)
SA	-1.08	0.07	3.14
	(0.34)	(0.15)	(1.45)
WA	-0.92	0.03	4.63
	(0.09)	(0.16)	(0.88)
TAS	-0.49	-0.10	4.13
	(0.20)	(0.12)	(0.73)

note: figures in brackets are standard errors.

It was pointed out above that a rise in income would probably not induce an individual consumer to raise his utilization of already owned durables but might encourage him to buy new durables and thereby raise his derived demand for electricity as an input. This seems to be borne out by the results here. The estimated income elasticity drops considerably (its highest value is now 0.17, in Victoria) and income does not appear to be a significant determinant of electricity demand in any State at all, since none of the income elasticities exceeds its standard error. It might be argued therefore that the use of a time trend to represent durable-related demand for electricity has been so important that it has completely driven out the influence of the income variable. This certainly seems to suggest that hypotheses that include assumptions about the influence of durable goods on electricity demand are more powerful than, say, the static demand model used previously.

It would appear also that in the static model the downward trend of relative electricity prices was picking up some of this "durables" effect. Electricity price still seems to have an important influence on electricity demand since 4 out of the 6 price elasticities exceed their standard errors, but the sensitivity to price changes has been considerably reduced with, now, only one elasticity being greater than 1. Interestingly enough, incorporating the rough measure of durables demand has made price more influential in Tasmania's equation, where the elasticity now has the expected negative sign and is more than twice its standard error.

In every case, the exponential trend reflecting holdings of durables is highly significant. It is not however easy to compare the different estimates of r, the rate of growth of durables holdings, between different States, because, as was explained above, the sample periods used differ significantly between different states. We can however compare the three states with the longest sample period, New South Wales, Victoria, and South Australia. The first two show relatively high growth rates of durable holdings per capita over the 1950's and 1960's, of 6.45% and 5.64% respectively, while South Australia has barely more than half these rates at 3.14%. It would seem clear that some (probably social) influences were at work in South Australia to differentiate it from its eastern neighbours, and several hypotheses may be put forward about these. However, more light may be shed on this feature if we can compare this model across all States. To do this means sacrificing many degrees of freedom to bring each State's sample period down to that of Queensland, which is 1962-3 to 1970-1. The results of these regressions are compared in Table X 5. Along with these results are the results from fitting the static model in its logarithmic version over the same sample period so that the effect of including a trend for durables holdings may be observed.

It now becomes apparent that South Australia rather than Tasmania is the "odd man out" in electricity demand over the 1960's. Leaving South Australia on one side, we may discuss the other results first. The income effect of course becomes noticeably less important when the static model is extended to become the trend model to allow for the influence of consumer durables. Income elasticities are considerably

Table X 5. Static Demand Model and demand model with time trend; common sample period 1962-3 to 1970-1.

STATE	PRICE ELASTICITIES		INCOME ELASTICITIES		r%
	STATIC	TREND	STATIC	TREND	TREND MODEL
NSW	-0.99	0.43	0.76	0.38	6.53
	(0.31)	(0.37)	(0.30)	(0.16)	(1.42)
VIC	-0.90	0.39	1.03	0.40	4.88
	(0.80)	(0.34)	(0.52)	(0.25)	(1.28)
QSLD	-1.42	-0.27	0.97	-0.07	5.85
	(0.35)	(0.24)	(0.32)	(0.19)	(1.25)
SA	-1.53	-1.46	0.14	0.26	-0.38
	(0.13)	(0.67)	(0.16)	(0.25)	(3.2)
WA	-1.34	0.26	0.68	0.08	9.6
	(0.71)	(0.42)	(0.65)	(0.21)	(1.9)
TAS	-1.34	-1.18	0.25	-0.19	2.40
	(0.37)	(0.26)	(0.19)	(0.16)	(1.10)

reduced, going from positive to negative in the cases of Queensland and Tasmania, and also lose some statistical significance when compared with their standard errors. Only in the case of South Australia does the income elasticity increase. The effect of introducing the trend variable also reduces the effect of falling electricity prices over the sample period. In the static model own price elasticity is everywhere negative and nearly one or greater. When allowance is made for the influence of consumer durables, however, the elasticities frequently become positive and little different from their standard errors suggesting again that price was picking up some reverse trend effect in the static model. Nevertheless in Tasmania and South Australia price remains an important influence on electricity demand. It would appear in the case of Tasmania that influences that differentiated consumer behaviour from that in other states in the longer sample period have relatively disappeared in the 1960's.

The estimates of the rate of growth or per capita consumer durable holdings are however the most interesting feature of Table X 5. New South Wales, Victoria and Queensland have relatively high growth rates of durable holdings. However Western Australia has the outstanding growth rate of 9.6% suggesting that this State really "took off" in terms of personal consumption in the 1960's. The build up of mining exploration and consequent settlement and housebuilding has been reflected in significant growth of holding of consumer durables. Tasmania, on the other hand, still exhibits some of the characteristics of a sluggish market with growth of only 2.4%.

The exceptional results are those of South Australia. In this State demand becomes more sensitive to income changes after allowance is made for the influence of consumer durable holdings, and at the same time, price remains just as significant a determinant of electricity demand: in both models the elasticity is more than twice its standard error. Even when the trend is explicitly accounted for, own price elasticity in South Australia is relatively high at -1.46, suggesting falling tariffs have been very important determinants of electricity demand movements over the 1960's. The estimated growth rate of durable holdings is the oddest feature of the South Australian results since it is negative though small and several times smaller than its standard error. An initial hypothesis to explain these results could be that South Australia, like Tasmania is a relatively sluggish market.

We inserted a trend variable to pick up the influence of durable holdings and at the same time noted that the price variable itself exhibited a declining trend over the sample period. It might be arguable therefore, that the price and income variables are picking up the influence of consumer durables holdings instead of the crude representation that apparently works in other states. This suggests the necessity of incorporating consumer durable influences on electricity demand into our equations in some more sophisticated manner than the crude approximations used here, and we turn to this task below. However, first of all, a few more comments may be made about the model under discussion. In discussing equation (11) above it was pointed out that it implies an unit elasticity of electricity demand with respect to durable holdings. Suppose however this elasticity was not 1 but, plausibly, less than one. Then the estimate of r in the fitted equations would be biased downwards since it would be in fact an estimate of (d.r) where d is the elasticity of electricity demand with respect to durable holdings. Hence a low elasticity figure, d, may account for the estimate of r being not significantly different from zero in the case of South Australia. Again, this failure to identify a parameter of the model suggests the need for a more complex model to pick up the influences we are discussing.

Finally a point made by Fisher and Kaysen may be recalled. They hypothesized that more urbanized states would show <u>greater</u> elasticities of income and price in a demand model like this, because rural populations would show a greater concentration of major electrical appliances in their holdings of electrical durables than urban populations because of the opportunity for substitution provided by local laundries and supermarket freezers in urbanized communities.

To a very small extent this is borne out in the data: New South Wales and Victoria both show higher income elasticities than Queensland, a state less urbanized.

NOTES TO CHAPTER X

1. The two leading academic investigations of residential energy demand that have been published, and to which further references are made, are: Fisher and Kaysen (1962) and Balestra (1967).

2. See the annual report for 1970/71 of the State Electricity Commission of Victoria.

3. In conversation with the marketing department of one of the largest Australian utilities (SECV), it was pointed out that "interruptibility" facilities are only rarely used in Australia and that effectively all customers, residential and industrial, have to take tariffs as given.

4. Balestra (1967).

5. Houthakker (1950).

6. This point is made by Prest (1963).

7. This study of residential demand forming Chapter X of the thesis was completed before the collection and analysis of the detailed tariff data reported in part I, Chapters II to V. It was felt that it was reasonable to leave it as a self contained analysis. However the chapter following (Chapter XI) does make use of the later tariff data in analysing demand for electricity, this time in the industrial sector.

8. Both Fisher and Kaysen (1962) and Balestra (1967) use this model to represent short run demand for energy usage.

9. See Fisher and Kaysen (1962).

CHAPTER XI

STOCK ADJUSTMENT MODELS OF RESIDENTIAL DEMAND

INCREASING THE SAMPLE SIZE

It has been obvious in the preceding chapter that the demand models used are too crude to make much sense of the electricity demand situation especially in its relationship with the demand for consumer durables. However these crude models are worth discussing because they can be applied at the state level without severe damage to the degrees of freedom in their estimation. (Economic models at the individual state level are regrettably few in Australia.) Nevertheless we have now come to the point where, in considering more complex models of demand, we are compelled to change over to demand equations fitted at the national level with pooled data from time series and cross section samples.

It was especially noticeable in the case of Tasmania that the equations in the previous chapter performed sensibly and could be compared well in the period of the 1960's. As a result, the pooled sample used from this point on consists of 10 annual observations on each State for the period 1962-62 to 1970-71, with the exception of Queensland. The sample size is therefore 50 and the observation matrix is X

$$X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ X^{(5)} \end{bmatrix}$$

where $X^{(i)}$ is the matrix of annual observations in the $i^{\underline{th}}$ state.

We could have chosen a slightly smaller size of sample and included Queensland but it seems to make slightly more sense, in using large numbers of variables, to maximize the degrees of freedom. For this reason the sample of 50 observations on 5 States over 10 years is preferred.

STOCK ADJUSTMENT

Since we now have a sample size that is more than adequate, an attempt to formulate some relatively more complex models can be made. In developing a research strategy to do this it was impossible to ignore the body of literature on stock adjustment models. For this reason it was decided to experiment with such models as much as possible before attempting any other analyses. Stock adjustment models that have appeared in the literature are stringent in the requirements put on the estimating equation and their main drawback is that they impose an all or nothing choice on the investigator. Either the precise form fits well or the whole rationale of the model has to be re-thought. However their impact has been such that it was essential at least to report experiments with them.

We investigate two closely related stock adjustment models here that have been successfully used in studies in the U.S. The first model (Houthakker and Taylor (1970)) was developed to be applied to every category of consumption demand reported in U.S. Statistics; in other words it shows the most general assumptions possible about consumer behaviour over a whole range of products. The second model, (Balestra (1967)) was developed in an attempt to apply the Houthakker-Taylor estimating equation to a particular energy market, but incorporating a slightly different rationalization of the estimating equation.

Both models rely heavily on special estimation techniques, for several reasons. The estimating equation can easily contain over-identified parameters when it is extended to include new variables and Houthakker and Taylor consequently used an interative estimating technique (three pass least squares), to identify the structural parameters they were particularly interested in. Balestra, besides tackling the identification problem by iteration, hypothesized that the fact that a pooled sample was being used was having an important influence on the specification of the error term, and used a two-stage procedure to obtain Generalized Least Squares estimators of his structural coefficients. The problems involved will become clear as we proceed. We begin by outlining the models used.

The estimating equation finally arrived at looks fairly familiar in stock adjustment models:

$$Q_t = a_0 + a_1 Q_{t-1} + a_2 \Delta X_t + a_3 X_{t-1} + v_t \qquad \ldots\ldots \quad (1)$$

where X_t is a vector of observations on exogeneous variables at time t, and v_t is an error term.

The generalized Houthakker-Taylor model is rationalized in these terms

$$Q_t = a + bS_t + cX_t \qquad \ldots\ldots \quad (2)$$

Equation (2) says demand for a commodity depends on some physical or psychological stock, S_t, and some exogeneous variables, e.g. personal income, price etc. S_t is a very general variable in the model and may represent nothing more than the habit-formed level of Q_t. In terms of fuel demand we would obviously like to interpret it as the habit-formed level of service flows from a stock of consumer durables.

In turn

$$\dot{S}_t = Q_t - dS_t \qquad \ldots\ldots \quad (3)$$

Or simply the rate of change of the stock variable is new demand less depreciation. It is rather difficult to conceive of a depreciation of psychological stock or habit formation; we can really only conceive of a gradual shedding of old preferences and an adoption of new ones. Thus the drawback of the Houthakker-Taylor model is that it is so general that its particular application to any one commodity raises conceptual problems. In any case, using (2) and (3) and the fact that

$$\dot{S}_t = \frac{1}{b}\left[\dot{Q}_t - c\dot{X}_t\right] \qquad \text{from (2),}$$

gives a first order differential equation, which can be approximated as a difference equation if we define

$$\dot{Q}_t = Q_t - Q_{t-1}$$

to yield the estimating equation (1)[1]. The essence of this model is that once a habit-formed psychological stock of consumption has been built up we can use the depreciation concept to obtain an estimating equation that contains only observable predetermined variables besides the observed demand. In our case the obvious exogeneous variables to use in equation (1) are price and income so that our estimating equation is:

$$Q_t = a_0 + a_1 Q_{t-1} + a_2 \Delta P_t + a_3 P_{t-1} + a_4 Y_t + a_5 Y_{t-1} + u_t \quad \ldots\ldots \quad (4)$$

The work of Balestra (1967) tries to apply this general equation to the fuel market using a more sensible rationalization, which proceeds as follows, (see also Balestra and Nerlove, 1966).

The starting point is to relate the demand for energy (E_t) to the stock of appliances (S_t) by some utilization rate (k)

$$E_t \equiv kS_t \quad \ldots\ldots\ldots\ldots \quad (5)$$

If r of these appliances wear out each year then $(1-r)S_{t-1}$ is the stock of appliances available in year t to require fuel demand of

$$k(1-r)S_{t-1} \quad \ldots\ldots\ldots\ldots \quad (6)$$

In period t however, actual fuel demand will not only include this demand level but also any new demand that has arisen over the period (t-1 to t). Actual fuel demand will be given by equation (5) so that the <u>new</u> fuel demand will be

$$E_t^* \equiv kS_t - k(1-r)S_{t-1} \quad \ldots\ldots\ldots\ldots \quad (7)$$

Substituting this equation into equation (6)

$$E_t^* \equiv E_t - E_{t-1} + rE_{t-1} \qquad \text{.......... (9)}$$

Thus we started with the notion of relating demand to a stock of appliances and have arrived at a relationship between new demand and demand levels previously established, in the process eliminating the level of appliance stocks. There will be a similar relationship for each particular form of energy used, including electricity

$$Q_t^* \equiv Q_t - Q_{t-1} + r_Q Q_{t-1} \qquad \text{.......... (10)}$$

where r_Q is the depreciation rate of electricity using appliances.

It now remains to set up the behaviour patterns determining demand, and to relate these to the identities we have just established. The Balestra model is quite simple in this respect:

new electricity demand is determined by new energy demand and the price of electricity

$$Q_t^* = b_0 + b_1 P_t + b_2 E_t^* \qquad \text{.......... (11)}$$

and actual energy demand depends on a vector of exogeneous variables, X_t

$$E_t = c_0 + c_1 X_t \qquad \text{.......... (12)}$$

Then, combining equations (9), (10), (11), and (12) we obtain the estimating equation[2]

$$Q_t = a_0 + a_1 P_t + a_2 \Delta X_t + a_3 X_{t-1} + a_4 Q_{t-1} + v_t \qquad \text{.....} \quad (13)$$

where v_t is an error term.

The only difference between (13) and (4) is that own price does not enter with a lag. Now, Balestra hypothesizes that the main element in the vector of exogeneous variables will be personal income accompanied perhaps by population. Hence the estimating equation will be

$$Q_t = a_0 + a_1 P_t + a_2 \Delta Y_t + a_3 Y_{t-1} + a_4 Q_{t-1} + v_t \qquad \text{.....} \quad (14)$$

To sum up the form of stock adjustment model used, the equations fitted are: two forms of the Houthakker-Taylor equation (4), with price and without price, and the Balestra and Nerlove equation (14).

The important question now is what do the regression estimates tell us about the underlying stock adjustment model? If in particular, we adopt the Balestra rationalization of the stock adjustment equation, we can begin to obtain estimates of the depreciation rates of electrical appliances in particular and of household energy using appliances in general.

THE SAMPLE

It has already been pointed out that the data consist of a pooled sample of cross section and time series observations, and there is a possibility of problems of non-homogeneity of the data arising. Characteristics of electricity demand in a state like Western Australia can be expected to differ from those of a state like Tasmania. Almost certainly if there are interstate differences they will be largely related to differing weather patterns. These weather patterns may cause demand in one state to differ by a given level from demand in another state.

There are several ways of tackling this situation and two were adopted in this study. Firstly we could add to the regression equations a variable to represent weather trends over the states and the period. It may be recalled that in the individual state regressions no weather variable was used. The objective was to investigate the explanatory power and plausibility of various elasticity estimates, and nothing was lost by considering the randomly distributed weather patterns as components of the disturbance term. However in the pooled sample approach it becomes advisable to recognize that explicit recognition of the weather variable may be necessary. The weather variable used is mean maximum temperature observed in the period July 1st to June 30th in each state in each year calculated from monthly data provided by the Commonwealth Meteorological Information Service.

An alternative strategy for dealing with interstate differences is to use a set of state-specific dummy variables. With this approach we can investigate whether the regression intercept terms differ significantly across states. The procedure for carrying out such tests is described in Johnston (1972). Essentially an analysis of co-variance test is carried out to determine whether the use of separate state intercepts adds significantly to the "explanation" of the overall variance.

The results of these two approaches - using a weather variable and using state dummies - is discussed in the presentation of the results below.

RESULTS FOR STOCK ADJUSTMENT MODELS

The first stock adjustment model estimated here is the one described by Balestra and derived above as equation (14). The results are first reported without the inclusion of the weather variable:

$$\underset{}{Q_t} = \underset{(113.35)}{58.78} + \underset{(0.03)}{1.01Q_{t-1}} + \underset{(11.99)}{12.72\,\Delta Y_t} + \underset{(3.40)}{1.76Y_{t-1}} - \underset{(0.34)}{0.22P_t}$$

$$R^2 = 0.99 \qquad \text{d.f.} \quad 45$$

The fit of this equation is not particularly good in terms of the standard errors of the coefficients. Recalling the derivation of equation (14), we can estimate depreciation coefficients from the regression coefficients on Y_{t-1}, ΔY_t and Q_{t-1}. The implied depreciation rate for all fuel using durables is 0.14, which does not seem unreasonable, though is possibly rather high. However, the implied depreciation rate for electricity using consumer durables is negative at -0.99 which does not make any kind of sense at all. Clearly the Q_{t-1} term is picking more effects than it should.

If we now include State dummy variables to try to pick out non-homogeneity over states the equation becomes:

$$Q_t = \underset{(275.58)}{52.25} + \underset{(0.14)}{0.92Q_{t-1}} + \underset{(15.01)}{16.06\Delta Y_t} + \underset{(17.05)}{11.22Y_t} - \underset{(0.70)}{0.37P_t}$$

$$-34.21(DVIC) -5.72(DSA) \quad -24.30(DWA) \quad .72.39(DTAS)$$

$$R^2 = 0.99 \qquad\qquad d.f. = 41$$

Where (DVIC), (DSA), (DWA), and (DTAS) are dummy variables for the states of Victoria, South Australia, West Australia and Tasmania. The depreciation rates have now reversed. The plausible estimate is that for electricity using durables, 0.08, but the depreciation rate for all fuel using durables is 0.69 which is improbably high. The other interesting coefficient in this model is that of long run price elasticity which can be estimated from the equations as:

model without dummies: -0.04

model with dummies : -0.06

Clearly the stock adjustment model suggests electricity demand is much less price sensitive than the naive models used earlier. When the naive model using price and income as the only independent variables is run for the pooled sample with state dummies, the estimated price elasticity is -0.21.

The use of the state dummies had the effect of making the estimates of the underlying structural coefficients slightly more plausible but it did not clearly improve the overall fit. The latter can be tested as was said above on an analysis of variance test. The reduction in unexplained variance by assuming different intercepts yields a calculated F value of 0.45 when compared with the regression that does not have different intercepts. On comparison with the tabulated F distribution, the hypothesis that different states have different intercepts is rejected. Hence non-homogeneity of the data - at least on this test - does not appear to be a serious problem.

The alternative method of allowing for interstate differences was by inclusion of a weather variable. The estimated equation with W_t as the weather variable was:

$$Q_t = \underset{(236.61)}{-63.56} + \underset{(0.04)}{1.03}Q_{t-1} + \underset{(12.10)}{12.38}\Delta Y_t + \underset{(3.72)}{0.92}Y_{t-1} - \underset{(0.34)}{0.22}P_t + \underset{(288.53)}{170.33}W_t$$

$$R^2 = 0.99 \qquad \text{d.f.} = 44$$

The weather variable is rejected as insignificant so again we have not found a difference between states in electricity demand. However the coefficient on Q_{t-1} remains implausible. Recalling that it was plausible when state dummies were included (even though these dummies were statistically insignificant taken together), the equation with weather and state dummies was tested. The results were:

$$Q_t = \underset{(617.04)}{404.16} + \underset{(0.13)}{0.93Q_{t-1}} + \underset{(15.69)}{13.39\Delta Y_t} + \underset{(17.18)}{10.67Y_{t-1}} + \underset{(0.71)}{0.32P_t} + \underset{(790.25)}{504.79W_t}$$

$$-51.74(DVIC) - 9.28(DSA) - 10.07(DWA) - 20.9(DTAS)$$

$$R^2 = 0.99 \qquad d.f. = 40$$

The depreciation and elasticity coefficients are then estimated as:

	depreciation rate on all fuel using durables	depreciation rate on electrical durables	price elasticity of demand
model without dummies	0.07	-0.97	-0.04
model with dummies	0.79	0.07	-0.06

As might be expected from the negligible effect of weather on the overall explanation the results are virtually unchanged, from those of the model without the weather variable.

This suggests one interesting though negative result; weather differences between states do not appear to explain the apparent differences in responsiveness of electricity demand to changes in economic variables.

The position that we have now reached is that there appears to be no significant interstate differences in the sample but only when state dummies are present do both the depreciation coefficients show the expected positive signs. Consequently a compromise equation was adopted. The estimates of r_Q - the depreciation rate of electricity using durables - was 0.07 in the equation just estimated. If this is taken as an estimate of the true r_Q we could derive the following constrained estimating equation:

$$Q_t - 0.93\, Q_{t-1} = a_0 + a_1 \Delta Y_t + a_2\, Y_{t-1} + a_3\, P_t + v_t$$

where v_t is an error term.

This is equivalent to estimating the basic equation subject to the restraint that $r_Q = 0.07$. (See Johnston 1972, pages 155-159.)

The results of running this equation are:

without a weather variable:

$$Q_t - 0.93\, Q_{t-1} = \underset{(64.22)}{330.16} + \underset{(12.83)}{15.02}\Delta Y_t + \underset{(3.62)}{0.51}\, Y_{t-1} - \underset{(0.17)}{1.08}\, P_t$$

$R^2 = 0.50$ d.f. = 46

and with a weather variable:

$$Q_t - 0.93Q_{t-1} = \underset{(122.23)}{459.45} + \underset{(12.76)}{14.93\Delta Y_t} + \underset{(3.89)}{2.33Y_{t-1}} - \underset{(0.26)}{0.83P_t} - \underset{(235.88)}{292.41W_t}$$

$R^2 = 0.52$ d.f. = 45

These results certainly seem to be a substantial improvement, in terms of plausibility of the coefficients and in terms of their relationship to their standard errors. (The reported R^2's are not comparable with those for the unconstrained equations.)

The implied depreciation coefficients are then:

electrical durables:	0.07 (a)	0.07 (b)
all fuel using durables:	0.04 (a)	0.17 (b)

where (a) refers to the equation without a weather variable and (b) refers to the equation with a weather variable. Since weather has made some contribution to the explanation (its regression coefficient exceeds its standard error) it seems more sensible to adopt this second version of the constrained stock adjustment model.

It appears from the second equation that price is still an important determinant of electricity consumption in the residential sector, (the regression coefficient is more than three times its standard error); the estimated long run price elasticity of demand is -0.15, which does not seem implausible on a prioti grounds. This elasticity estimate suggests however that while price has been significant in explaining movements in electricity demand, non optimal pricing policies that set price in excess of or below marginal cost may not have caused very large losses in the consumers' surplus of residential electricity consumers.

Care has been taken in this chapter to investigate fully the application of a stock adjustment model to residential demand. Such models could not be ignored because they have played a significant part in the econometrics of demand. However, it is only with difficulty that anything like satisfactory results have been achieved, and it therefore seems worthwhile turning attention away from such model forms where the particular form of the estimating equation is of such importance to more general models of demand. This is done in the next chapter; we take with us the results that interstate differences and the influence of different weather patterns are both of negligible significance in explaining demand in the residential sector.

NOTES TO CHAPTER XI

1. Proceed as follows:

$$Q_t = a + bS_t + cX_t \qquad \ldots\ldots (2)$$

$$\dot{S}_t = Q_t - dS_t \qquad \ldots\ldots (3)$$

Then differentiating with respect to time:

from (2)

$$\dot{Q}_t = b\dot{S}_t + c\dot{X}_t \qquad \ldots\ldots (2a)$$

and from (3)

$$\dot{S}_t = Q_t - \frac{d}{b}\left[Q_t - a - cX_t\right] \qquad \ldots\ldots (3a)$$

Solving for $\dot{Q}_t$ from (2a) and (3a)

$$\dot{Q}_t = (b-d)Q_t + ad + cdX_t + c\dot{X}_t \qquad \ldots\ldots (3b)$$

We now convert to discrete time functions by setting

$$\dot{Q}_t \simeq \Delta Q_t = Q_t - Q_{t-1}$$

and $\dot{X}_t \simeq \Delta X_t = X_t - X_{t-1}$

Then:

$$Q_t = (b-d)Q_t + ad + Q_{t-1} + c\Delta X_t + cd\ (\Delta X_t + X_{t-1})$$

giving

$$Q_t = \frac{ad}{1-(b-d)} + \frac{1}{1-(b-d)}Q_{t-1} + \frac{c(1+d)}{1-(b-d)}\Delta X_t + \frac{cd}{1-(b-d)}X_{t-1}$$

which simplified is:

$$Q_t = a_0 + a_1 Q_{t-1} + a_2 \Delta X_t + a_3 X_{t-1}$$

2. The derivation of equation (13) is the following:

From equation (10)

$$Q_t = Q_t^* + (1-r_Q) Q_{t-1}$$

$$\therefore \quad Q_t = b_0 + b_1 P_t + b_2 E_t^* + (1-r_Q) Q_{t-1} \quad \text{from equation (11)}$$

$$\therefore \quad Q_t = b_0 + b_1 P_t + b_2 (\Delta E_t + rE_{t-1}) + (1-r_Q) Q_{t-1} \quad \text{from equation (9)}$$

$$\therefore \quad Q_t = b_0 + b_1 P_t + b_2 c_1 \Delta X_t + b_2 c_1 r X_{t-1} + (1-r_Q) Q_{t-1} \quad \text{from equation (12)}$$

which can be rewritten:

$$Q_t = a_0 + a_1 P_t + a_2 \Delta X_t + a_3 X_{t-1} + a_4 Q_{t-1} + v_t$$

where v_t is an error term; this is equation (13)

Depreciation coefficients can be identified from the restrictions

$$a_2 = b_2 c_1 \; ; \quad a_3 = b_2 r c_1 \; ; \quad a_4 = (1-r_Q)$$

CHAPTER XII

A GENERALISED MODEL OF RESIDENTIAL DEMAND

INTRODUCTION

Not being entirely satisfied with the stock adjustment model we discussed the possibility of fitting a more general model of demand. One model which has stood the test of time in demand studies is the partial adjustment model. This would simply relate electricity demand to its own past value and a series of exogenous variables which are thought to determine the demand for the flow of electricity services

$$Q_t = f\,(Q_{t-1},\ X_{1t}\ X_{2t}\ \ldots)$$

This sort of model has a much less specialised form than the stock adjustment model previously discussed and allows the researcher to experiment substantially with the explanatory variables X_1 X_2 ... etc. Several experiments of this sort are reported here. However the next section - which is in the form of a digression - presents a model of how an estimating equation like this could be derived from an explicit consideration of the simultaneous determination of the demand for the services of a stock of durables and the demand for electricity as an input to provide these services. Such a model seemed to be the essential next step from naive demand models and a first attempt at it is found in the stock adjustment models of the previous chapter.

DEMAND FOR CONSUMER DURABLE SERVICES

The model to be discussed here is an attempt to consider explicitly the simultaneous determination of the demand for the services of a stock of consumer durables (d) and the demand for electricity as an input to provide those services.

We can describe the services of a stock of durables at the time t as c(t) and the utility from those services as $u\left[c(t)\right]$. Since durables have usually a significant life the consumer's planning horizon will be relatively long, so that his maximand will be the discounted value of the utility he desires from his durable holdings:

$$\int_o^T e^{-rt} u\left[c(t)\right] dt \qquad \ldots\ldots\ldots (1)$$

where r is the consumer's rate of discount. However the costs of obtaining these flows of services are the expenditure on durables and expenditure on electricity input: $p_d d + p_q q$. If we assume these terms reflect completely the cost involved in using his stock of durables then the consumer maximizes

$$\int_o^T e^{-rt}\left[u\left[c(t)\right] - p_d d - p_q q\right] dt \qquad \ldots\ldots\ldots (2)$$

One restraint on this maximand is the limit to services which a stock of durables can provide when combined with electricity input, i.e. the services production function:

$$f(c,q,d,) = 0 \qquad \ldots\ldots\ldots (3)$$

Finally, the depreciation of the stock of durables should be considered. If h is new purchases of durables then:

$$h = \dot{d} + gd \qquad \ldots\ldots\ldots (4)$$

where g is the rate at which durables depreciate. If now equation (2) is maximized subject to the restraints embodied in equations (3) and (4) (a problem in the calculus of variations) we obtain as one of the consumer's equilibrium conditions:

$$\frac{u'(c)}{p_q} = \frac{dq}{dc} \qquad \ldots\ldots\ldots (5)$$

which simply states that the ratio of the marginal utility of services' consumption to the price of electricity equals the rate at which electricity input increases when services consumption increases. In other words the consumer equates the marginal utility form extra consumption of services to their marginal cost.

We may put very general formulations in the functions involved here, and two obvious ones are:

$$u\left[c(t)\right] = \frac{1}{1-v}\, c(t)^{1-v} \qquad \ldots\ldots\ldots (6)$$

and:

$$c = b_o q^{b1} d^{b2} \qquad \ldots\ldots\ldots (7)$$

(6) is simply a constant elasticity of marginal utility function to represent the consumer's preferences; it has several useful properties in consumer theory, e.g. positive but diminishing marginal utility of consumption, and constant elasticity of marginal utility with respect to additional consumption:

i.e. $u'(c) = 1/c^v$

$u'(c) \rightarrow \infty$ as $c \rightarrow 0$

$u''(c).\ c/u = 1/1-v$

(7) is merely the Cobb-Douglas formulation of the services production function. Using (6) and (7) in the above analysis allows us to write (5) as:

$$q = \frac{a}{P_q} c^{1-v} \qquad \text{........ (8)}$$

where a is the elasticity of output of durable goods services with respect to electricity input.

Using (8) as a model allows us to incorporate the demand for the services of durable goods <u>directly</u> into the demand for electricity when we interpret c as the usage of or demand for the services of a set of consumer durables. We need now an equation to determine c.

It can be expected that a variety of social demographic and economic factors will determine the desired flow of services of a stock of consumer durables and quite a degree of experimentation is carried out here. If these variables are represented by the vector $z_1 \ldots z_n$, the estimating equation becomes:

$$q^* = a_o p_q^{a_1} z_1^{a_2} z_2^{a_3} \ldots z_n^{a_{n+1}} v \qquad \ldots\ldots\ldots (9)$$

where v is some error term, and q* is desired electricity consumption rather than actual. This can be used to allow for some lag in the operation of maximizing utility, so that

$$\left[\frac{q^*_t}{q_{t-1}}\right]^w = \frac{q_t}{q_{t-1}} \qquad \ldots\ldots\ldots (10)$$

Writing primes for logarithms the estimating equation then becomes

$$q'_t = wa'_o + wa_1 p'_{q_t} + wa_2 z'_{1_t} + \ldots + wa_{n+1} z'_{nt} + (1-w)q'_{t-1} + u_t$$

$$\ldots\ldots\ldots (11)$$

and this is the equation that is experimented with.

Each structural coefficient is identified as:

$$a_i = \frac{m_{i-1}}{(1-m_{n+1})}$$

where m_i is the i^{th} particular regression coefficient and m_{n+1} is the coefficient on q'_{t-1}.

SPECIFYING THE DEMAND FOR ELECTRICITY

The introduction indicated that a general model to explain the demand for electricity as derived from the demand for consumer durable services could be formulated. Having seen in the above section one way in which such a model might be rationalised, we now return to the main theme and consider what variables might be used to explain the demand for the services of consumer durables.

Without attempting a detailed social-psychological approach, it nevertheless makes sense to include some long term social variables that will affect the demand for consumer durables. As each new generation becomes independent, new family units are set up and consequently an impetus is given to the demand for consumer durable services, especially with the well known tendency for the setting up of nuclear rather than extended kinship family units. Hence it makes sense to include the variable, marriage registrations (M), to pick out this effect. However, new demand for consumer durables often arises when an existing family unit moves house and from this point of view some measurement of new house ownership is required. The variable used here is Completions of All New Homes and Flats (A). Almost certainly this will be collinear with M, even when normalized by population changes so we may only be able to pick out sensible effects from one of these variables. Both of these variables can be expected to influence the long term equilibrium ownership of consumer durables.

Before demand reaches this point however, it can be expected that economic factors will be influential in determining consumer durable demand and hence derived electricity demand. Since we have begun to expand on the familiar partial adjustment model by including social influences on durable demand, it also makes sense to expand the list of likely economic influences. Price and income of course are obvious candidates. Given the current state of published statistics, we are compelled to use instrumental variables in many cases.

As an instrumental variable for prices of electricity using consumer durables the published Price Index of Household Supplies and Equipment (PD) is used.

So far for estimation of income effects studies including this one have used current real personal income. However the usefulness of a measure of permanent income is well known in applied econometric studies and it is used here. The measure is a weighted moving average of past incomes with the weights decaying geometrically. With a given truncation point, n, current permanent income Y_t^P is:

$$Y_t^P = \frac{1}{\sum_{i=0}^{n} \beta^i} \sum_{i=0}^{n} \beta^i Y_{t-i} = 0.1632 \sum_{i=0}^{7} (0.9)^i Y_{t-i}$$

when β is chosen as 0.9 and n is chosen as 8. (These parameters were suggested by the Reserve Bank of Australia's macro-econometric model's consumption equations; see Norton, 1970.)

However, it is apparent that while permanent income makes a sensible budgetary constraint, it does not exhaust the sources of funds for purchase of durables. Hence as an estimate of additional funds available, another variable is used: Instalment Credit balances outstanding with Retail Business (RCR).

Having described the variables used to measure social and economic influences, it, finally, seems wise to take account of another well known factor in consumer durable demand - the demonstration effect.[(1)]

The demonstration effect suggests that the desire for services of a new durable arises from contact with economic units already owning the durable. To use Fisher & Kaysen's analogy the growth of durable ownership is like the spread of a contagious disease. For this reason we could pay attention to the variable: (stock of consumer durables at beginning of period t) as determining the demand for services of electricity using durables during period t and hence the derived demand for electricity. The closest we can come to an estimate of stocks of electrical goods is to build up an estimate of stocks of all durables. The method used here is that outlined by Stone and Rowe (1957) which is briefly described as follows.

New purchases of consumer durables at time t, C_t, can be regarded as the sum of replacement demand and new additions to stock, U_t and V_t respectively:

$$C_t = U_t + V_t$$

If we assume that the amount of stock used up approximates declining balance depreciation at a rate of 1/N per period, and if we assume a smaller proportion 1/M of new purchases depreciates per period, we have

$$U_t = \frac{S_t}{N} + \frac{C_t}{M}$$

where S_t is the stock at the beginning of period t. If we now define the stock at the beginning of period t + 1 as the sum of stock at the beginning of period t and net additions to stock during the period t we have

$$S_{t+1} = S_t + C_t - U_t$$

By substituting in the previous equations we then obtain

$$S_{t+1} = \frac{N-1}{N} S_t + \frac{M-1}{M} C_t$$

In other words stock at the beginning of period t + 1 is a linear function of stock at the beginning of period t and consumption during period t, with weights (N-1/N) and (M-1/M) which are both functions of the depreciation rate 1/N:

$$\frac{1}{M} = 1 - 1/N \log_e \left(\frac{N}{N-1}\right)$$

A set of data on stock of consumer durables is then built up using the equation for S_{t+1} in the following manner. An initial value (stock at the beginning of 1959) is obtained from a Reserve Bank Study for Australia as a whole[(2)] and allocated to each State on the basis of the contemporaneous population. Five different values of N are then chosen (N = 2, 3, 4, 5, 6) and the equation estimated. (Actually estimated for financial year estimates of the starting value, obtained by linear interpolation.) These five series for each State can then be used in estimating an equation like (9) or (11) and effectively, by choosing the series which yields overall maximum R^2 we obtain maximum likelihood estimates of our structural parameters[(3)].

To sum up therefore the variables thought to contribute to movements in the demand for the services of consumer durables are a sub-set of:

(i)	$(PD)_t$	=	price of consumer durables at time t
(ii)	Y_t^P	=	permanent income
(iii)	S_t	=	stock of consumer durables at the beginning of time period t
(iv)	M_t	=	marriage registrations
(v)	A_t	=	completions of all houses and flats
(vi)	RCR_t	=	credit balances outstanding with retail businesses in real terms
(vii)	P_t	=	price of residential electricity consumption

THE DEMONSTRATION EFFECT

We begin this part of the estimation by considering what values of S_t we could use. The method I have chosen for this is to run separate regressions for each State using the simple equation:

$$Q_t = b_0 + b_1 Y_t^P + b_2 \frac{S_t}{(PD)_t}$$

The optimal series of S_t could then be chosen for each State from the values of the R^2 and standard error of estimate. These Stock values might then be used in the pooled sample estimation of this model, to capture the demonstration effect. The results of the experiment are tabulated in Table XII, 1. They are rather surprising. The iterations do not seem to have converged as successfully as might be hoped on a maximum value of R^2. Only in the case of Victoria and West Australia does a maximum value for R^2 stand out. In the other states, no clear maximum is obtained, and several turning points occur. The most interesting result is that it is the shortest estimate of durable life that does best in explaining electricity demand. (Stone and Rowe found similar corner maxima at the lowest value of N when they applied a form of this model to British data, Stone and Rowe (1960).) This suggests that, if the stock of durables is indeed giving rise to a form of demonstration effect in determining the derived demand for electricity, it is the more recent purchases that are having the greatest influence; in other words the stock series that does best in 3 out of 5 cases is the one which gives highest weighting to its most recent components, or writes off new items of stock most quickly. Now this is not at all implausible.

TABLE XII, 1

Maximum Likelihood Estimation of the depreciation coefficient in Consumer durable stocks.

		NSW (R^2)	VIC (R^2)	SA (R^2)	WA (R^2)	TAS (R^2)
N = 2	Y = 3	0.9933*	0.9945*	0.9755*	0.9975*	0.9218
N = 3	Y = 6	0.9927	0.9939	0.9762	0.9973	0.9150
N = 4	Y = 8	0.9906	0.9915	0.9710	0.9915	0.9266
N = 5	Y = 10	0.9913	0.9827	0.9843	0.9826	0.9544*
N = 6	Y = 13	0.9915	0.9784	0.9914*	0.9751	0.9441

* = maximum

NOTE: N is the reciprocal of the fraction of Stock written off in one year; Y is the number of years taken to write off 90% of the initial value of a unit of stock, when 1/N is the depreciation coefficient.

It makes sense to say that it is the most recent (and therefore most in accord with current fashion or taste) acquisitions of consumer durables that have the greatest demonstration effect. On this basis, it can be argued that a series on recent acquisitions alone will give the best representation of the demonstration effect; in addition, the failure of the iterative estimation to converge within the reasonable interval of durable life (3-13 years), suggests replacing the stock variable with a measure of recent acquisitions. Hence in the regressions which follow the variable used to capture the demonstration effect is the deflated value of personal consumption expenditure on household durables, lagged one period $(PCED/PD)_{t-1}$.

RESULTS

The results from experimenting with this formulation of the demand for electricity in the residential sector are shown in the large table, XII, 3. Before looking at individual results in detail, the overall specification of the model can be examined by looking at table XII, 2. In this table the estimates of W, the rate of flow adjustment parameter, and the estimates of the exponent of the price term in the optimizing condition $U'(c) = p\ (dq/dc)$ are shown. It may be recalled that in looking at one theoretical basis of the estimating equation, the exponent of the price term turned out to be unity (see the second section of this chapter). If this overall specification is to make any sense at all we clearly require W lies between zero and unit and that the exponent of p be approximately unity. We may recall lagged consumption persistently entered with a coefficient greater than unity in the estimation of the stock adjustment models, until some restrictions

TABLE XII, 2

Regression No.	Estimate of W	Estimate of Exponent on price term
1	0.08	1.3
2	0.08	1.3
3	L.15	1.1
4	0.11	1.0
5	0.16	1.5
6	0.19	1.3
7	0.14	1.3
8	0.14	1.2
9	0.05	
10	0.16	1.5
11	0.24	1.2
12	0.09	1.3
13	0.08	1.3
14	0.15	1.3
15	0.09	1.3
16	0.07	1.0
17	0.08	1.5
18	0.08	1.2
19	0.08	1.2
20	0.07	1.3

TABLE XII, 3

No.	Constant	Q_{t-1}	P_t	$(PD)_t$	$(PCED/PD)_{t-1}$	Y^P_t	M_t	A_t	$(RCR)_t$	T_t	$\bar{R}^2$	d.f.
1	1.20	0.92	-0.11	0.27	0.004	-0.04					0.99442	44
	(0.98)	(0.03)	(0.08)	(0.20)	(0.01)	(0.09)						
2	1.02	0.92	-0.10	0.23		-0.006					0.99454	45
	(0.87)	(0.03)	(0.07)	(0.17)		(0.05)						
3	3.36	0.85	-0.17	0.57	0.02	-0.11				-0.34	0.99463	43
	(1.64)	(0.05)	(0.09)	(0.27)	(0.01)	(0.10)				(0.21)		
4	1.73	0.89	-0.11	0.32		0.009				-0.16	0.99453	44
	(1.14)	(0.04)	(0.07)	(0.20)		(0.05)				(0.16)		
5	2.40	0.84	-0.23	0.24	0.01	0.06			0.03		0.99463	43
	(1.23)	(0.06)	(0.11)	(0.19)	(0.01)	(0.11)			(0.02)			
6	3.62	0.81	-0.24	0.47	0.02	-0.02			0.02	-0.25	0.99462	42
	(1.65)	(0.06)	(0.11)	(0.29)	(0.01)	(0.13)			(0.02)	(0.23)		
7	1.78	0.86	-0.18	0.17		0.10			0.03		0.99465	44
	(1.03)	(0.05)	(0.09)	(0.18)		(0.09)			(0.02)			

TABLE XII, 3 (cont'd)

No.	Constant	Q_{t-1}	P_t	$(PD)_t$	$(PCED/PD)_{t-1}$	Y^P_t	M_t	A_t	$(RCR)_t$	T_t	$\bar{R}^2$	d.f.
8	1.95	0.86	−0.17	0.21		0.09			0.02	−0.06	0.99452	43
	(1.16)	(0.05)	(0.10)	(0.22)		(0.10)			(0.02)	(0.19)		
9	0.59	0.95				−0.09	0.20	−0.02			0.99454	45
	(0.23)	(0.02)				(0.06)	(0.09)	(0.03)				
10	3.21	0.84	−0.23	−0.03	0.02	−0.09	0.23	0.007	0.04		0.99461	41
	(1.41)	(0.06)	(0.12)	(0.27)	(0.01)	(0.15)	(0.15)	(0.03)	(0.02)			
11	7.74	0.76	−0.29	0.30	0.06	−0.47	0.43	0.05	0.03	−0.68	0.99522	40
	(2.27)	(0.06)	(0.11)	(0.28)	(0.02)	(0.20)	(0.17)	(0.03)	(0.02)	(0.27)		
12	1.58	0.91	−0.10			0.03					0.99445	46
	(0.77)	(0.03)	(0.07)			(0.04)						
13	1.45	0.92	−0.11		−0.002	0.05					0.99432	45
	(0.98)	(0.03)	(0.08)		(0.009)	(0.07)						
14	2.29	0.85	−0.21			0.15			0.03		0.99467	45
	(0.87)	(0.05)	(0.09)			(0.08)			(0.02)			

TABLE XII, 3 (cont'd)

No.	Constant	Q_{t-1}	P_t	$(PD)_t$	$(PCED/PD)_{t-1}$	Y^P_t	M_t	A_t	$(RCR)_t$	T_t	$\bar{R}^2$	d.f.
15	0.60	0.91	-0.12			0.04		-0.003			0.99432	45
	(0.81)	(0.04)	(0.07)			(0.05)		(0.03)				
16	1.19	0.93	-0.07			-0.06	0.14				0.99454	45
	(0.81)	(0.03)	(0.08)			(0.07)	(0.09)					
17	1.76	0.92	-0.12		0.003						0.99434	46
	(0.86)	(0.03)	(0.08)		(0.005)							
18	1.54	0.92	-0.10						0.0002		0.99434	46
	(0.80)	(0.03)	(0.07)						(0.009)			
19	1.39	0.92	-0.10				0.07				0.99456	46
	(0.76)	(0.03)	(0.07)				(0.05)					
20	1.48	0.93	-0.10					0.009			0.99434	46
	(0.79)	(0.03)	(0.07)					(0.02)				

were put on the estimating equation. This problem does not arise with the model estimated here. Nevertheless we cannot be entirely satisfied with the estimates of the rate of adjustment. In most cases they suggest that between 10% and 30% of the difference between last year's consumption and this year's desired consumption, in logarithmic terms, is made up by the difference between last year's consumption and this year's actual consumption. This is a very slow rate of flow adjustment, though more plausible a priori than the rates estimated in the stock adjustment models discussed previously. The estimated exponent on the price term also lies acceptably close to its a priori value and overall these results suggest that investigating the demand for electricity by a direct utility maximizing specification rather than the more indirect familiar stock adjustment models, results in equations that behave more realistically, at least in working with this type of pooled sample.

However, one obvious problem to be expected when this direct approach is made to explain determinants of the demand for consumer durable services is that severe multicollinearity in the specified explanatory variables will obscure their individual effects. There is evidence in several of the equations of a problem of multicollinearity. One particularly odd set of results is associated with the permanent income variable. This often has the wrong sign and is not significantly different from zero when there are fairly large numbers of other explanatory variables included. At the same time, in the first ten or twelve regressions the variable $(PD)_t$ relative price of household supplies and equipment, also performs very oddly, consistently having a positive sign. It is apparent from looking at the raw data that durable prices have been rising in step with permanent income over the 1960's. Consequently the only reasonable estimating approach is to discard one of the variables causing the multicollinearity and the sensible one to discard is $(PD)_t$ since it consistently was the wrong a priori sign. The results presented in Table XII, 3 begin by showing equations that are quite large in the number of explanatory variables and proved by a process of elimination to develop a sensible estimating equation that picks up a workable approximation to the behavioural specification derived above.

INDIVIDUAL VARIABLES

We may examine the performance of the individual variables in turn. Lagged residential electricity consumption, Q_{t-1}, and the constant term are important in the degree of explanation achieved. The estimated rate of adjustment is fairly low, as already pointed out, but this seems to be a frequent result in estimating this type of equation by ordinary least squares.

The price variable, P_t, relative average price of electricity, remains as important as ever, its regression coefficient always exceeds its standard error by a large amount and we have already concluded that its true exponent in the a priori behavioural relation has been estimated sensibly. This result in particular is a welcome sign that this approach to estimation makes sense, and that the general formulations imposed on the model are not too restrictive. The short run elasticity of demand with respect to price (ignoring, for the moment the restrictions on the dynamic model), as estimated by the partial regression coefficient in these logarithmic regressions, varies between -0.07 and -0.20. Thus, although price is a significant determinant of demand, the short run elasticity is quite low. It is not possible to be dogmatic about the implications of this for electricity pricing policy. Roughly we know that, in consumers' surplus terms, the higher the own price elasticity, the greater the welfare loss from setting price in excess of marginal cost. While no attempt is made at this point to quantify this welfare loss, we may say that if the bottom range of the estimate is accurate, the welfare loss is very small in partial equilibrium consumers' surplus terms. On the other hand, the equations that seem to be most acceptable on a variety of grounds, (discussed below), provide estimates of the short run elasticity that lie in the top part of the range, say about -0.2; however, this is still a relatively low figure. At this point therefore it may be suggested that, in the residential sector at any rate, non-optimal pricing policies would not be likely to involve great welfare loss, at least as estimated by this method.

As was pointed out above, the variable $(PD)_t$ representing consumer durable prices and measured by the index of household supplies and equipment, performed very oddly wherever it was included. It only had the correct sign once, in regression no. 10, and was very insignificant in that case. Hence it was dropped from regressions 12 to 20, which resulted in an improvement in the performance of the permanent income variable, Y_t^p.

It was also pointed out above that the best way to capture the demonstration effect on demand for consumer durables resulting from their increased ownership throughout the economy seemed to be to use the most recent acquisition of durables as an explanatory variable. Real personal consumption expenditure on durables lagged one period, $(PCED/PD)_{t-1}$ was used for this purpose. It seemed more plausible not to deflate it by population since what was wanted was a measure of new durable ownership in the economy as a whole. On the whole this variable performed reasonably well, (it only once had the wrong sign and was twice its standard error on 4 out of the remaining 7 occasions) but was also associated with poor performance of the permanent income variable. This latter result does not make a great deal of sense since, on the whole, we might feel happier with an equation in which permanent income influenced the demand for durables rather than relying on an estimated demonstration effect. Nevertheless, recent expenditure on durables does appear to influence the demand for electricity, and this must really be a form of demonstration effect. Superficially, we could argue that demand for electricity must go up when people have more durables simply because of the production function for the output of durables services. But recent expenditure on durables will not pick up this production effect unless it happens that newly acquired durables are (a) a very large part of the stock of all durables owned and (b) new durables are less efficient in using electricity than older durables. Since both these assumptions are implausible, we are left with the conclusion that some type of demonstration effect is operating.

It has already been pointed out that the performance of the permanent income variable, Y_t^p, was very sensitive to the other explanatory variables included in the estimating equation. However when a smaller model is estimated, (leaving out $(PD)_t$ and $(PCED/PD)_{t-1}$) the performance of the permanent income variable makes some sense. Its overall impact, however, remains small, and this makes reasonable sense. We saw in an earlier chapter that explaining electricity consumption by income, though a fairly obvious approach, nevertheless requires careful reationalization; it is clearly naive to hypothesize that electricity consumption by the average consumer rises just because his income rises. Rather, it makes more sense to envisage permanent (or even transitory) income as representing the position of the consumer's budget restraint in his durable purchases, and this seems to be the effect picked up in this model. Electricity demand, as formulated here, is subject to a flow adjustment process operating when electricity is viewed as the input into the production of a flow of durable services. Consequently Q_{t-1} will be a very significant variable and the influence of Y_t^p will only be felt when it is used as a principal determinant of the flow of durable services, in its function as part of the consumers' budget restraint. This hypothesis makes for a realistic rationalization of the performance of Y_t^p in determining electricity demand, especially when it appears that the specification estimated here works more plausibly than most of the models in the literature.

The other variable included to pick up the effect on the demand for durable services of shifts in the consumer's budget restraint is $(RCR)_t$, real per capita borrowings from hire purchase organizations. We need to be very careful about using a variable of this nature. In a macro-consumption function a variable of this sort is sometimes included as a wealth or liquidity constraint, with the obvious drawback that credit conditions are as endogenous to a macro model as consumption itself. However it can be argued here that the partial equilibrium nature of the model makes the use of a variable like $(RCR)_t$ permissible. The performance of the variable is quite reasonable. It always has the expected positive sign, and exceeds its standard error in 6 out of the 8 cases used. It is apparent therefore that movements in the consumer's budget constraint are important determinants of demand for durable services and therefore for the flow of electricity input.

Several "social" or "demographic" variables were included with the specific purpose of making the model as general as possible. The "social" variables used here are M_t and A_t; respectively, the number of registered marriages per head of the population and the number of new dwelling units constructed per head of the population. A model emphasising social variables was run as equation 9, but did not give plausible results. The $\bar{R}^2$ was comparable with other formulations but the individual regression coefficients on Y_t^p and A_t were unacceptable and an extremely low rate of flow adjustment was implied by the coefficient on Q_{t-1}. Re-running this regression with individual state dummies added nothing whatsoever. Nevertheless one social variable, M_t, performed quite well when used in conjunction with economic variables in a form of mixed model. To the extent that new family units move into existing accommodation rather than new dwellings, it is clear that M_t will have greater influence on the rising demand for durable services than A_t.

Weather was experimented with as a variable (T_t) but performed so badly in having wrong signs and insignificant coefficients that it was discarded in the later regressions.

This completes discussion of the effects of individual variables. It remains now to consider overall equation performance. The coefficient of multiple determination corrected for degrees of freedom ($\bar{R}^2$), the general significance of individual coefficients and their a priori plausibility make up three reasonable grounds for choosing between different equations, although it has already been pointed out, evidence of multicollinearity limits the choice of explanatory variables that prove acceptable. For this reason equation, No. 11, though it has the highest $\bar{R}^2$ cannot be adopted. The variables Y_t^p and $(PD)_t$ do not work well together. Nevertheless this equation does show the highest rate of flow adjustment observed, 0.24.

The next best equation on the grounds of highest $\bar{R}^2$ is equation No. 14 and this has much to recommend it. Besides achieving high $\bar{R}^2$, all the regression coefficients are well in excess of their standard errors and the estimated rate of flow adjustment is in the top part of the range of estimates. This model contains only "economic" variables: the variables determining the demand for durable services are permanent income and real per capita credit balances. This model suggests that the only important determinants of the demand for durables are the variables in the consumer's budget restraint; in other words there is a large unsatisfied market for durables and any outward shifts in the average consumer's budget plane allow him to raise immediately his purchases of durables, and hence his purchases of electricity. This is a significant result for electricity authorities in particular and fuel suppliers in general. It suggests that there is room for rapid expansion in durable sales to the average consumer in Australia as a whole whenever the level of aggregate demand and flow of credit are expanded. None of the other specifications work as well as this one, which suggests this purely economic flow adjustment model is the best choice for understanding the determinants of residential electricity consumption.

The equations that are only just less satisfactory than this one suggest similar results. Equation 7 is identical to equation 14 except for the inclusion of the variable $(PD)_t$, which has the wrong sign but is not significantly different from zero. Equations 5 and 6 lose precision in the estimates of the coefficients on $(PD)_t$ and Y_t^p but $(RCR)_t$ still has the right sign, though a lower degree of significance. Overall, therefore, economic variables seem most important in determining durable demand and wherever social variables are used they add nothing or even detract from the overall performance of the equation. The economic variables of prime importance are those representing the budget restraint of the consumer; relative price of durables and the demonstration effect do not perform nearly so well in considering the overall reasonableness of the equations.

NOTES TO CHAPTER XII

1. For some British evidence on this, see Williams (1972).

2. See the comments in Norton and Broadbent (1970).

3. For a proof, see Nerlove (1960).

CHAPTER XIII

TESTING FOR PRICE INFLUENCES IN INDUSTRIAL AND COMMERCIAL ELECTRICITY DEMAND

This closing chapter in the empirical demand section has a limited purpose. As already explained, the main empirical work on demand - for his residential sector - was the initial part of the work for this study. That work, reported in the previous chapters, reviewed and estimated a variety of electricity demand models in the usual approach of the econometrics literature.

However of great interest to economists is the attempt to measure price elasticities. In the course of preparing this thesis data on actual tariffs were collected (as reported in Part I), and as a final exercise it seems reasonable to try to take advantage of this data in looking at the demand in two sectors that were not analysed previously - the industrial and commercial sectors.

The purpose of this chapter therefore is simply this: to investigate the price elasticity of demand using the conventional average price measures and to contrast this with estimates of price elasticity using the newly collected data on actual tariff rates. The researcher may ask himself:

(a) on what basis should the price of electricity be calculated?

(b) what differences will show up in the economically important estimates of price elasticity of demand?

(c) how are these differences to be evaluated and what lessons are to be learned from the exercise?

PREVIOUS EXERCISES

As has already been seen many of the previous electricity demand studies adopted the device of a single average price for electricity in the particular sector in question. The average price is simply the average revenue associated with kilowatt hour sales.

There is no doubt that this is a convenient approach and faced with the multiplicity of published electricity prices many researchers have considered that it is the only possible approach. In the industrial and commercial sector itself, previous studies - which are rare - have used this approach, e.g. Baxter and Rees (1968)[1] and Bell (1973).

Nevertheless, the early classic electricity demand study of Houthakker[2] rejected the use of average revenue as an adequate measure of the price of electricity, and it is worth considering the arguments for and against this practice.

There is no obstacle to using average revenue when only one figure is given as "the price of electricity" in the tariff schedule itself. The problems arise when some form of two part or multi-part tariff is used. Revenue in a simple two part tariff is:

$$R = A + pQ$$

where A is the standing charge, p the variable charge and Q is sales volume. Average revenue is then

$$\frac{R}{Q} = (p + \frac{A}{Q})$$

which is negatively related to the quantity purchased, (Q).

The effect of this can be illustrated in a simple demand equation. Suppose quantity demanded is related to an index of final production and to price. Using average revenue as a measure of price gives:

$$Q^D = \alpha + \beta_1 X + \beta_2 (p + A/Q) + u$$

where X is final production and u is a disturbance term.

If this equation is fitted several consequences may result:

(i) by hypothesis Q and u are related, but Q is used as an explanatory variable as well as being the dependent variable. What we have is a special case of the disturbance term not being independent of the explanatory variables. It is well known that inconsistent estimates of the regression coefficients may result; (see Goldberger 1964, page 278). This is the sort of problem that arises in simultaneous equation models when single equation estimation techniques are used.

Usually we would hope to avoid the problem by formulating some sort of simultaneous equation model. However in the sense that our faulty equation arises from a problem of the data and its measurement, this way out is not really available to us. The argument of Houthakker who first raised the point was that only "marginal price" data will do. Average revenue data by itself is economically meaningless.

(ii) The situation may be even worse than this. Rather than a meaningless correlation we may obtain a fitted equation that performs well, theoretically and statistically but is in fact misleading. By definition there must be a negative relationship between

$$Q \quad \text{and} \quad (p + \frac{A}{Q})$$

and this may persuade us that demand is more negatively related to price than is the case.

There therefore seem to be two constraints on estimating these demand functions. If average revenue and true marginal price are compared the following results might be expected:

(i) If using "marginal price" really is an improvement in the data matrix, then the overall degree of explanation ought to improve. Improving our data will give more efficient estimators resulting in a lower overall standard error of estimate, a higher $\bar{R}^2$ - the degree of explanation obtained.

Thus if there is a significant negative relationship between price and quantity demanded, using "marginal price" data should improve the estimation of this relationship.

(ii) On the other hand however, if the effect of using average revenue is not to pick up imperfectly a significant relationship, but rather to impose a spurious relationship, using marginal price data will result in a much poorer overall explanation and may raise the possibility of wrong (i.e. positive) signs on the price variable.

With these factors in mind, several experiments were carried out to compare average revenue and marginal price relationships with quantity demanded.

DATA

Average revenue is of course easy to calculate; obtaining a "marginal price" for electricity is much harder, given the multiplicity of available tariffs. Only a partial solution is available to the problem. For different states we can think of many tariff series which are candidates for "marginal price". After some careful consideration two or three were isolated. These were large block discount rates chosen to correspond as nearly as possible with the theoretical ideal of the expenditure on the marginal unit of electricity.

Using the coding introduced in Part I of the thesis, the representative marginal prices were:

Victoria:	CBV4	-	fourth block discount for commercial use
	BV3	-	third block discount for industrial use.
South Australia:	ES8	-	eighth block discount for all customers
	RBS2	-	second block discount for day usage in industry.
West Australia:	BW4	-	fourth block discount for industrial power.
Tasmania:	CBT3	-	third block discount for commercial use
	BT4	-	fourth block discount for industrial use.

Initially we make the assumption that these rates represent the price of electricity for commercial and industrial use <u>at the margin</u> of choice. This assumption is discussed again in the conclusions to this chapter.

<u>DEMAND EQUATIONS</u>

We now want to specify a hypothetically reasonable demand function for commercial and industrial use. This function has to be largely free of specification errors but as general as possible since we want to compare results for the same equation simply using different measures of "price": i.e. average revenue or "marginal price" as represented by the actual tariff rates already chosen above.

The form adopted reflects (a) the idea that demand for electricity input depends on the demand for final output and relative prices and (b) that there is a lagged adjustment between actual demand and equilibrium demand. This gives us a simple adjustment equation:

$$Q_t = \alpha + \beta_1 I_t + \beta_2 (RP)_t + \beta_3 Q_{t-1}$$

where Q_t = demand for electricity in year t

Q_{t-1} = demand for electricity in year t-1

I_t = the ANZ bank index of industrial production for all states in year t

RP_t = relative price of electricity in year t; this is either

(a) average commercial or industrial revenue deflated by a general index of prices

or (b) marginal tariff rate for commercial or industrial demand, again deflated by a general index of prices.

This equation is identical to the formulation that Baxter and Rees[3] found to fit electricity demand "best" in terms of overall degree of explanation. We have fitted the equation in logarithmic form so that the regression coefficients directly measure the constant elasticities of demand with respect to the different explanatory variables.

For each category of demand reported: separate industrial and commercial demand for Victoria, South Australia and Tasmania, and total industrial and commercial demand for West Australia, the equation is fitted twice. Once average revenue is used to calculate price and once a marginal tariff rate is used to measure price. The equations can then be compared on standard errors of each regression coefficient and on overall explanation: $\bar{R}^2$, the coefficient of determination corrected for degrees of freedom is used for the latter.

RESULTS

The simple formulation performed very well in total. None of the equations exhibited any indication of serial correlation of the residuals on the usual tests.

The detailed results are presented in Table XIII,1 below, and we can organize our comments about them under several headings.

Overall explanation is high even for Tasmania which shows up the poorest results. As expected industrial production has always a positive impact on electricity demand, and in most cases the elasticity of demand with respect to final production is less than one[4]. The lagged dependent often improves the explanation but is by no means a dominant influence, and in several cases is unsignificant - particularly in the Tasmanian results.

TABLE XIII,1. Industrial and Commercial Demand in four states; relative price tests.

Regression coefficients and standard errors of explanatory variables.

	Dependent Variable	Industrial Production	Average Revenue	Marginal Price	Lagged Dependent	$\bar{R}^2$
1a	Victorian Industrial	0.48	-0.49		0.33	0.98
		(0.19)	(0.25)		(0.11)	
1b	Victorian Industrial	0.65		0.12	0.36	0.96
		(0.34)		(0.38)	(0.17)	
2a	Victorian Commercial	0.45	-0.25		0.69	0.99
		(0.50)	(0.48)		(0.32)	
2b	Victorian Commercial	0.47		-0.093	0.74	0.99
		(0.64)		(0.45)	(0.38	
3a	South Australian Industrial	0.28	-1.79		-0.04	0.99
		(0.43)	(0.28)		(0.14)	
3b	South Australian Industrial	1.58		-2.37	-0.43	0.98
		(0.43)		(0.47)	(0.20)	

	Dependent Variables	Industrial Production	Average Revenue	Marginal Price	Lagged dependent	$\bar{R}^2$
4a	South Australian Commercial	0.83	−0.83		0.28	0.99
		(0.30)	(0.42)		(0.13)	
4b	South Australian Commercial	1.10		0.84	0.64	0.99
		(0.25)		(0.40)	(0.12)	
5a	West Australian Industrial and Commercial	0.92	−0.69		0.43	0.99
		(0.17)	(0.19)		(0.08)	
5b	West Australian Industrial and Commercial	1.04		−0.54	0.50	0.99
		(0.28)		(0.41)	(0.13)	
6a	Tasmanian Industrial	1.59	−0.39		0.022	0.67
		(0.78)	(1.27)		(0.44)	
6b	Tasmanian Industrial	1.63		0.47	−0.05	0.68
		(0.77)		(1.10)	(0.47)	
7a	Tasmanian Commercial	1.60	−0.97		−0.87	0.74
		(0.95)	(0.27)		(0.67)	
7b	Tasmanian Commercial	0.68		−2.35	−0.13	0.64
		(1.19)		(0.82)	(0.83)	

Our main interest however is directed towards the price variable. When average revenue is used as a measure of price, the elasticity of demand is always negative and in four cases out of seven is clearly significant at the five per cent level. The elasticities estimated this way are usually less than unity, with the exception of the case of South Australian industrial demand. In this latter case the average revenue variable appears to have swamped the influence of the lagged dependent.

On this basis we are tempted to report significant negative relationships between price and quantity demanded. However examining the use of marginal tariff rates to represent price we see a very different picture. To begin with the overall degree of explanation usually falls, though the change is often very marginal. In only one case does the precision of the estimated equation improve when marginal price rather than average revenue is used; (this is the case for Tasmanian industrial demand). It looks therefore as if using "marginal price" does not amount to an improvement in the data, since we cannot detect an overall better fit for the equations. Our estimates become less rather than more efficient.

Does this mean that average revenue data introduces a spurious correlation which disappears when marginal tariff rates are used instead? This is indeed arguable. The marginal price variable shows a negative relationship with demand in only four out of the seven cases, and in only one of those cases (Tasmanian commercial demand) is the elasticity statistically significant.

The usual effect of using marginal price seems to have been to throw more weight on the lagged dependent variable as an explanatory influence on demand. In one case this is even associated with a significant positive relationship between marginal price and demand.

The overall impression of the results therefore seems to be:

(a) using marginal price does not result in a more precise fit of an already established price-demand relationship.

(b) rather it is the case that using average revenue seems to introduce a negative relationship that disappears when marginal tariff rates are used to measure price.

CONCLUSIONS

The overall impression of the results is clear enough. In this concluding section however several counter arguments on the problems of testing can be suggested.

It was suggested that marginal price was the relevant economic factor influencing demand - rather than average revenue. Against this, it has often been suggested that complex pricing structures make it extremely difficult for consumers to calculate marginal price. Turvey has often argued for simplicity in tariff schedules[5]. If there are information costs in calculating the price of electricity at the margin, what is the consumer to use as a rule of thumb? It may be that average revenue (or average bills) is the indicator of electricity price that the consumer uses. Hawkins (1973) suggests that the consumer may use such a measure as a less costly way of evaluating price.

This obviously raises difficulties for using the theory of demand to predict behaviour. If the consumer uses average revenue as an indicator of price he may falsely think marginal price is falling when simply sales volume is rising. The question is - does he therefore react by expanding demand so that the negative relationship between average revenue and demand is a true behavioural result rather than simply a spurious correlation

It is clear that a great deal of theoretical and empirical work needs to be carried out relating demand to the price which consumers think reflects choice at the margin. At the moment it is difficult to decide whether or not to question the conventional approach of using average revenue as a measure of price at the margin.

There are also more practical reasons for not using the results of Table XIII,1 to reject the conventional approach. Houthakker's original argument implies that all electricity pricing is on a two part basis. But this is certainly not the case. There is a huge variety in tariff types used. This could mean two things. Firstly the term

$$A/Q$$

which forms part of average revenue in a simple two part tariff may in practice be <u>negligible</u> in calculating an overall average revenue for a large electricity authority which uses a whole variety of tariff types from flat rates through maximum demand charges to peak and off peak differentials. It can conceivably be the case that even with multi-part tariffs, average revenue remains in practice a good indicator of the price of electricity. There is a corollary to this argument. To represent marginal price we had to use a selection of marginal tariff rates as "representatives" of the concept we are trying to measure. However these are by necessity imperfect representatives - not every consumer will qualify for the largest block discounts, and not every consumer will pay the representative tariff or perhaps even anything like it.

The counter arguments therefore can be summed up as:

(i) marginal price is difficult to calculate; consumers may use average revenue as their price indicator.

(ii) in a complex tariff system, there is no guarantee that the simple two part tariff example does not over-emphasise what could be a negligible influence from fixed charges.

(iii) representative marginal tariff rates may be themselves imperfect guides to choice at the margin.

What we have discussed in this chapter indicates the difficulties involved in demand estimation. The conventional approach - with its apparently successful results - has come under fire, but the criticisms themselves may be open to doubt.

NOTES TO CHAPTER XIII

1. The reference to Baxter and Rees (1968) is an implied one. This study is discussed below, but in this context it apparently (but did not explicitly state the fact) used average price of electricity measured by industrial average revenue. Bell (1973) repeats some of the Baxter and Rees experiments and explicitly states the use of average revenue data.

2. Houthakker (1950).

3. Baxter and Rees (1968). See reference to 1 above.

4. Bell (1973) has several comments to make on whether or not this result is to be expected.

5. e.g. Turvey (1971a) pages 32-42.

PART IV

CONCLUSION

CHAPTER XIV

SUMMING UP

This study has tried to examine some of the interesting economic questions that fall within the competence of an outside observer of the Australian electricity industry. These aspects are both theoretical and empirical, though investigation of actual behaviour has to be limited to those areas where published statistics are readily available.

PART I

The price mechanism and its working is of perennial interest to the economist and the actual pricing behaviour of various Australian electricity authorities is investigated in the first part of the thesis. The number and complexity of the pricing schemes in use is something that immediately strikes the observer; nevertheless it was shown (in chapter II) that there can be an economic or engineering rationale for most of the tariff structures adopted. What has to be tested is whether these tariff structures in practice achieve the objectives that they are designed for. The bulk of the evidence on tariff and capacity and running cost correlations presented in chapter III seems to suggest that the tariffs used in the four states in the sample were not adequately reflecting their underlying theoretical principles. Many of the standing charges were more closely correlated with running costs and many of the marginal energy charges seemed to be related to capacity cost movements. The overall impression was of a badly organized tariff structure in many cases covering all of the states in the sample.

Such a finding begs two questions at least: are significant costs involved in the failure to allocate resources optimally and is demand sensitive enough to respond to tariff re-structuring? Chapters IV and V attempted a beginning on the answers. Chapter IV tried to examine the influences determining unit costs in various categories for different electricity authorities. A good deal of attention was paid to load factor as a measure of the degree of peak spreading and an attempt was made to measure the cost savings of achieving higher load factors. The conventional econometric techniques used gave grounds for believing that rises in load factor well within the historically established capabilities of the electricity authorities could provide substantial cost savings to the authority (and, in passing, this was related to the controversial Lake Pedder Conservation issue).

Chapter V was concerned with modelling the determination of load factor on the basis of economic behaviour. The burdens borne by the available data were heavy but we observed the expected result that - to a certain extent - load factor variation could be attributed to the availability or otherwise of time of day tariff differentials. A simultaneous equation model gave an estimate of the elasticity of load factor with respect to the peak - off peak price ratio.

Very briefly, Part I could be said to have posed the questions:

(i) how have tariff principles been formed?

(ii) have the principles operated in practice?

(iii) does it matter in cost terms if tariff principles are abused?

(iv) can tariffs in practice affect demand variability?

and to have provided the following answers:

(i) in many ways

(ii) no

(iii) yes

(iv) yes

PART II

The second part of the study moved into much more abstract territory. Any observer of the economics of electricity cannot fail to take account of the enormous advances made in recent years in investment planning principles for electricity generation. This part of the thesis - the theoretical as opposed to the empirical aspect of the title - examined the nature and developments of investment planning models. In this area, the name of Ralph Turvey stands out and chapter VI presented a detailed study of one of his most famous models - that contained in his 'Economic Analysis and Public Enterprise' (Turvey, 1971a). This model has to be seen in the context of other Turvey models and frequent reference was made to his other writings. Chapter VI tried to bring out some of the more neglected aspects of the Turvey approach - for example the formal nature and problems of evaluating Kuhn-Tucker multipliers as marginal costs, and more importantly the extensive implications for optimal amortization rules and their relation to the pricing and investment rules which Turvey highlights.

However, Turvey is neither the first nor only economist to tackle capital theory in general or optimal investment and amortization in particular. Chapter VII examined a series of models and contributions - selected from a vast literature - which have most closely been related to the Turvey approach. Harold Hotelling's classic article on depreciation (Hotelling (1925)) was among the first and best of these. There have been similar attempts in the accounting literature which are surveyed and recent neoclassical investment theory is also examined. The overwhelming impression that emerges is that to ensure a model is dynamic, the variables must depend systematically on time through allowing for technological progress, time dependent capacity and replacement costs or time dependent demand requirements. Simply "dating" the variables and discounting does not make a model dynamic in the sense that one part of the chosen solution cannot be isolated from another. It is regrettable that so many so-called dynamic models of investment planning fail to be more than a sequence of static models, each quite separate from the others. Quite a lot of the neoclassical investment models are at fault here. This particular point is illustrated mathematically in the appendix to part II of the study. Chapter VII, in addition, discusses some work by W.A. Lewis which once more shows how this writer was able many years ago to anticipate so many of the modern advances in price theory and investment planning.

Chapter VIII is a slight digression from the theme of the theoretical discussion. It presents a practicable investment planning model for an Australian electricity authority using the Kuhn-Tucker programming analysis discussed previously. Rather than use a set of monthly, daily or even hourly output requirements as the basis of the model, it uses a yearly output programme measured in terms of two factors: an average annual committment to supply and a peak load demand occurring for part of the year. This allows us to develop clearer expressions for peak and off peak marginal costs and shadow prices than are provided by the more conventional model of chapter VI. Additional constraints are also considered in analysing the investment programme.

Chapter IX returns to an aspect of the model that also appears extensively in the earlier chapters: the calculation of marginal costs. This is not of course the usual text-book version; Turvey's analysis rejects this out of hand in a long run model allowing for technological progress. Previous chapters had already established the programming interpretation of marginal cost whereby it is calculated from the Kuhn-Tucker multiplier associated with a given output requirement constraint. In addition to this version, Turvey in his paper on "Marginal Cost" (Turvey (1969)), and his book on "Optimal Pricing and Investment in Electricity Supply" (Turvey (1968)) had suggested a much more revolutionary definition of marginal cost as the addition to total system costs from advancing by one time period the starting date of an already forecast permanent rise in output requirements. Turvey did not provide a specific way of calculating such a measure that related to the programming version. Chapter IX tries to do this - and in the process examines and amends a French model of investment planning which attempted to extend the marginal cost concept into dynamic economics, (Albouy and Nachtigal (1970)). The problem is to measure "historical dynamic marginal cost" - and Chapter IX tries to show several ways of doing this and obtains two measures: dynamic marginal cost and instantaneous dynamic marginal cost; the latter is identical to the revolutionary version of Turvey (1969). In this exercise several insights into criticisms of the Turvey model and into dynamic optimization and pricing are gained.

PART III

In part III of the study, the empirical side is taken up again with a consideration of the demand for electricity. The problems of data collection loom largest here, particularly since detailed economic data on incomes and spending patterns did not receive attention from the official statisticians in Australia until relatively recently.

Chapter X considers some naive though interesting demand models for the residential sector estimated at the state level. The outcome of this exercise is a realisation that residential demand for electricity has to take account of the demand for the services of a stock of electricity using durable goods. This presupposes the use of some rather more sophisticated and complex demand models. These in turn are demanding in their data requirements so that, to keep sample sizes relatively large, a pooled estimation approach has to be adopted.

Chapter XI uses pooled cross section and time series data to fit one particular model of demand that has had a significant place in the econometric literature. At first sight it seems to be a model that exactly fits our requirement of investigating simultaneously the demand for electricity and the demand for durable goods services. This is the dynamic stock adjustment model (Houthakker and Taylor (1970)). The results from using the model are not entirely satisfactory however, and similar difficulties with it seemed to have been found in natural gas demand analysis in the U.S. (Balestra 1967). Consequently Chapter XII tries to incorporate the demand for durable goods more directly into the demand for electricity with a partial adjustment model that tries to be as general as possible in its choice of explanatory variables. This attempt is relatively successful.

The bulk of the demand analysis has been concentrated on the residential sector. Chapter XIII therefore turns finally to the industrial and commercial sectors to investigate demand patterns. The particular objective is the study of price response. The work on residential demand used the conventional "average revenue" estimate for price in the demand equations, but there are reasons for believing this may be unsatisfactory in the complex tariff cases of electricity demand. Therefore the availability of detailed tariff data collected for part I of the thesis (undertaken some time after the residential demand study) presented an opportunity to use the industrial and commercial demand estimates to test price response in more detail. Chapter XIII discusses the results of fitting a standard demand equation with different sets of price data and the arguments that follow about measuring price response. It is found that the "marginal price" estimate produces a far lower or insignificant responsiveness of demand than a price measure based on average revenue. Because multipart tariffs mean that average revenue must by definition fall as sales rise it could be argued that the conventional measures of elasticity overestimate demand response and cannot be consistently estimated. However, there is an equally strong case for suggesting that the "marginal price" of electricity is a poor guide to how consumers measure the overall price of electricity, so that the controversy over measuring price elasticity is not easily resolved.

This concludes my theoretical and empirical study of the supply and demand for electricity in Australia.

BIBLIOGRAPHY

A. REPORTS AND STATISTICAL SOURCES

Department of National Development (1967) Energy in Australia Commonwealth Government Printer, Canberra.

Electricity Supply Association of Australia, Statistics of the Electric Supply Industry (1953-1972), Melbourne.

Electricity Commission of New South Wales, Annual Report, (various issues).

State Electricity Commission of Victoria, Annual Report, (various issues).

Queensland State Electricity Commission, Annual Report, (various issues).

Electricity Trust of South Australia, Annual Report (various issues).

State Electricity Commission of Western Australia, Annual Report (various issues).

The Hydro-Electric Commission, Tasmania, Annual Report (various issues).

Snowy Mountain Authority, Annual Report (various issues).

B. REFERENCES

Albouy, M. & Nachtigal, J. (1970) — Extension of the Marginal Cost Concept in Dynamic Economics. European Economic Review, Spring, p. 327-356.

Arrow, K.J. (1964) — Optimal Capital Policy, the Cost of Capital and Myopic Decision Rules, Annals of the Institute of Statistical Mathematics, 16, pp. 21-30.

Arrow, K.J. & Kurz, M. (1971) — Public Investment, the Rate of Return, and Optimal Fiscal Policy, John Hopkins Press, Baltimore.

Balestra, P. & Nerlove, M. (1966) — Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: the Demand for Natural Gas. Econometrica, 34, pp. 585-612.

Balestra, P. (1967) — The Demand for Natural Gas in the United States, North Holland, Amsterdam.

Balinski, M.L. & Baumol, W.J. (1968) — The Dual in Non-linear Programming and its Economic Interpretation. Review of Economic Studies, 35, pp. 237-256.

Baumol, W.J. & Bradford, D.F. (1970) — Optimal Departures from Marginal Cost Pricing. American Economic Review, 60, pp. 265-283.

Baxter, R. & Rees, R. (1968) — The Industrial Demand for Electricity, Economic Journal, 78, pp. 277-298.

Bell, A. (1973) — Industrial Electricity Consumption - An Example of an Intermediate Good, Journal of Industrial Economics, 21, pp. 95-109.

Boiteux, M. (1949) — La Tarification des Demandes en Point: Application de la Theorie de la Vente au Cout Marginal. Revue Generale de l'Electricite, 58, pp. 321-340; translated (1960) as Peak Load Pricing, Journal of Business, 33, pp. 157-179.

Durbin, J. (1970) — Testing for Serial Correlation in Least Squares when some of the Regressions are Lagged Dependent Variables, Econometrica, 38, pp. 410-421.

Dhrymes, P. & Kurz, M. (1964) — Technology and Scale in Electricity Generation. Econometrica, 32, pp. 287-315.

Eckstein, O. & Fromm, G. (1968) — The Price Equation, American Economic Review, 58, pp. 1159-1183.

Fisher, F. & Kaysen, C. (1962) — The Demand for Electricity in the United States, North Holland, Amsterdam.

Galatin, M. (1968) — Economies of Scale and Technological Change in Thermal Power Generation, North Holland, Amsterdam.

Godley, W. & Nordhaus, W. (1972) — Pricing in the Trade Cycle, Economic Journal, 82, pp. 853-882.

Gregory, R.G. & James, D.W. (1973) — Do New Factories Embody Best Practice Technology? Economic Journal, 83, pp. 1133-1155.

Goldberger, A. (1964) — Econometric Theory, John Wiley & Sons, New York.

Gould, J.P. (1968) — Adjustment Costs in the Theory of Investment of the Firm. Review of Economic Studies, 35, pp. 47-55.

Hall, R.E. (1968) — Technological Change and Capital from the point of Views of the Dual. Review of Economic Studies, 35, pp. 35-46.

Hawkins, R. (1974) — The Retail Electricity Market: A Study of the residential and commercial sectors in New South Wales and the Australian Capital Territory; Working Paper 21, Department of Economics, Research School of Social Studies, Australian National University.

Helliwell, J. (and others) (1969) — The Structure of RDX1, Bank of Canada Staff Research Studies, No. 3.

Henderson, A.M. (1947) — The Pricing of Public Utility Undertakings, The Manchester School of Economics and Social Studies, 15, pp. 223-250.

Hirshleifer, J. (1958) — Peak Loads and Efficient Pricing: Comment Quarterly Journal of Economics, 72, pp. 451-462.

Hotelling, H. (1925) — A General Mathematical Theory of Depreciation. Journal of the American Statistical Association, 20, pp. 340-353. (reprinted in Parker, R.H. and Harcourt, G.C. (1969) (eds.) Readings in the Concept & Measurement of Income, pp. 261-275, Cambridge University Press.

Houthakker, H. (1950) Some Calculations on Electricity Consumption in Great Britain. Journal of the Royal Statistical Society, 114, (A), pp. 359-371.

Houthakker, H. (1951) Electricity Tariffs in Theory and Practice. Economic Journal, 61, pp. 1-25.

Houthakker, H. & Taylor, L. (1970) Consumer Demand in the United States 1929-1970 Analysis and Projection, Harvard University Press. (Second edition).

Johansen, L. (1959) Substitution versus Fixed Production Coefficients in the Theory of Economic Growth: A synthesis. Econometrica, 27, pp. 157-175.

Johnston, J. (1961) Statistical Cost Analysis, McGraw-Hill, New York.

Johnston, J. (1972) Econometric Methods, McGraw-Hill, New York, (2nd Edition).

Jorgenson, D. (1965) Anticipations and Investment Behaviour in Duesenberry (et.al.), The Brookings Quarterly Econometric Model of the United States, Rand McNally, Chicago.

Jorgenson, D. (1966) The Embodiment Hypothesis, Journal of Political Economy, 74, pp. 1-17.

Kay, J.A. (1971) Recent Contributions to the Theory of Marginal Cost-Pricing; some comments. Economic Journal, 81, pp. 366-371.

Kendall, M. (1948) Rank Correlation Methods, Charles Griffin, London.

Komiya, R. (1962) Technological Progress and the Production Function in the United States Steam Power Industry. Review of Economics and Statistics, 44, pp. 156-166.

Lewis, W.A. (1941) The Two Part Tariff, Economica, 8, pp. 249-270.
and also The Two Part Tariff, a reply, Economica, 8, pp. 399-408.

Lewis, W.A. (1949) Fixed Costs, Chapter 1 of Overhead Costs, Allen & Unwin, London.

Lewis, W.A. (1960) Depreciation and Obsolescence as Factors in Costing, Chapter 1 of Depreciation & Replacement Policy, edited by L. Meij, North Holland, Amsterdam.

Lipsey, R. & Lancaster, K. (1956) The General Theory of the Second Best, Review of Economic Studies, 24, pp. 11-32.

Littlechild, S.C. (1970) Marginal Cost Pricing with Joint Costs, Economic Journal, 80, pp. 323-335.

Lucas, R.E. (1967) Adjustment Costs and the Theory of Supply, Journal of Political Economy, 75, pp. 321-334.

Massé, P. (1962) Optimal Investment Decisions, Prentice-Hall Inc. Englewood Cliffs, New Jersey.

Massé, P. & Gibrat, R. (1957) Application of Linear Programming to Investments in the Electric Power Industry, Management Science, 3, pp. 149-166.

McColl, G.D. (1972) Cost of Supplying Electricity in Australia, Australian Economic Papers, 11, pp. 65-88.

Meek, R. (1968) The New Bulk Supply Tariff for Electricity Economic Journal, 78, pp. 43-76.

Nerlove, M. (1960) The Market Demand for Durable Goods: a Comment. Econometrica, 28, pp. 132-142.

Nerlove, M. (1972) Lags in Economic Behaviour, Econometrica, 40, pp. 221-251.

Norton, W.E. (and others) (1970) A Model of the Australian Economy, Reserve Bank of Australia Occasional Paper No. 3A.

Norton, W.E. & Broadbent, J.R. (1970) Equations for Personal Consumption Expenditure, Reserve Bank of Australia, Occasional Paper No.3B.

Parmenter, B.R. & Webb, L.R. (1974) Amortization and Public Pricing Policies, unpublished research paper, Department of Economics, University of Melbourne.

Posner, M.V. (1973) Fuel Policy: A Study in applied economics, Macmillan, London.

Prest, W. (1963) The Electricity Supply Industry, Chapter 4 in A. Hunter (Ed.) The Economics of Australian Industry, Melbourne University Press.

Ramsey, F.P. (1928) A Mathematical Theory of Saving, Economic Journal, 38, pp. 543-559.

Rees, R. (1968) Second Best Rules for Public Enterprise Pricing, Economica, 35, pp. 260-273.

Rowe, D.A. & Stone, R. (1960) The Durability of Consumers' Goods, Econometrica, 28, pp. 407-416.

Salter, W.E.G. (1966) Productivity & Technical Change, Cambridge University Press (second edition).

Schulze, C. & Tryon, J. (1965) Prices & Wages in Duesenberry (et.al.) The Brookings Quarterly Econometric Model of the United States, Rand McNally, Chicago.

Solow, R.M. (1959) Investment and Technical Progress in Arrow, Karlin and Suppes, (Eds.) Mathematical Models in the Social Sciences, Standford University Press, California.

Soper, C.S. (1973) Amortization - a comment on the Desrousseau Rule Research Paper, No. 6. Department of Economics, University of Melbourne.

Steiner, P.O. (1957) Peak Loads and Efficient Pricing, Quarterly Journal of Economics, 71, pp. 585-610.

Stone, R. & Rowe, D.A. (1957) The Market Demand for Durable Goods, Econometrica, 25, pp. 423-443.

Theil, H. (1971) Principles of Economics, North Holland, Amsterdam.

Turvey, R. (1964) Marginal Cost Pricing in Practice, Economica, 31, pp. 426-432.

Turvey, R. (1968) Optimal Pricing and Investment in Electricity Supply, Allen & Unwin, London.

Turvey, R. (1969) Marginal Cost, Economic Journal, 79, pp. 282-299.

Turvey, R. (1971a) Economic Analysis and Public Enterprise, Allen & Unwin, London.

Turvey, R. (1971b) A Reply, Economic Journal, 81, pp. 371-375.

United Nations (1972) Electricity Costs & Tariffs, A General Study (ST/ECA/156), Department of Economic and Social Affairs, United Nations, New York.

Williams, R. (1972) Growth in Ownership of Consumer Durables in the United Kingdom, Economica, 39, pp. 60-69.

Williamson, O.E. (1966) Peak Load Pricing and Optimal Capacity under Indivisibility Constraints, American Economic Review, 56, pp. 810-827.

Wright, F.K. (1964) Towards a General Theory of Depreciation, Journal of Accounting Research, 2, pp. 80-90, reprinted in Parker and Harcourt (Eds.) 1969, Readings in the Concept and Measurement of Income, pp. 276-289, Cambridge University Press.

Wright, F.K. (1968) Measuring Asset Services: A linear programming approach, Journal of Accounting Research, 6, pp. 222-236.